Sensor-Based Feedback for Piano Pedagogy

Vom Fachbereich Informatik der
Technischen Universität Darmstadt genehmigte

Dissertation

zur Erlangung des akademischen Grades des

Doktor-Ingenieur (Dr.-Ing.)

von

Aristotelis Hadjakos
(Dipl.-Informatiker, Dipl.-Musiklehrer)
geboren in Frankfurt a. M.

Erstreferent:	Prof. Dr. Max Mühlhäuser
Zweitreferent:	Prof. Dr. Paul Lukowicz
Tag der Einreichung:	20. Mai 2011
Tag der Disputation:	04. Juli 2011

Darmstadt, 2011

D17

Abstract

Recent advances in sensor technology provide new opportunities for applications that utilize the user's movement as an input source. This thesis focuses on movement analysis, which is a sub-area of a larger field concerned with the interpretation of sensor signals of human movement. Movement analysis can be used to provide feedback for training motor tasks, which is of interest for application areas such as rehabilitation, sports, and ergonomics.

Consciously controllable, goal-directed movements, which we call primary movements, lead to slight movements in other parts of the body that are beyond conscious control (secondary movements). Secondary movements are generated due to the mechanical interaction with the environment and physiological dependencies of the body. This thesis contributes methods to distinguish between primary and secondary movements. This is necessary in order to provide high-quality feedback for motor tasks that show a significant amount of secondary movement. This can be the case when the secondary movements are large because of large reaction forces (e. g., when shooting a ball) or when the primary movements to execute the task are small (e. g., in various forms of handcraft or musical instrument performance). Furthermore, a precise distinction between primary and secondary movement can be necessary to check whether the user keeps a part of the body still (e. g., as required by a gymnastic exercise). Apart from sensor-based feedback, our results can also be used to improve current gesture recognition methods by ignoring secondary movement in the sensor signal to avoid that a secondary movement is misinterpreted as an execution of a gesture.

The effectiveness of the proposed methods is shown in the context of pianist arm movements, which are particularly challenging to analyze. The distinction between primary and secondary movement in one joint of the arm is based on the measured movement in that particular joint, an estimation of key reaction force from MIDI data, and the movement in the other joints of the arm. In order to know on which arm the estimated key reaction force acts, it is necessary to determine which hand has played a note. For that purpose two methods are introduced: One method is based on MIDI; the other one uses data from inertial sensors in combination to MIDI. A third method based on Computer Vision, which was originally developed for sign language recognition, is evaluated here for tracking pianist hands.

Based on the analysis methods, two pedagogical applications were developed: One application supports an existing piano pedagogical movement notation and checks whether the player's movement conforms to the indicated movement. A user study with piano students of a music university shows that potential users judge that the system is useful for the training of technique. The second application visualizes the sensor data and allows synchronizing different performances of the same piece, making it easy to spot differences where a closer examination may be beneficial.

Zusammenfassung

Fortschritte im Bereich der Sensortechnik ermöglichen heute neue Anwendungen, bei denen die Bewegungen des Nutzers als Eingabequelle dient. Das zentrale Thema dieser Arbeit ist die Bewegungsanalyse. Sie ist Teil eines größeren Forschungsfeldes, nämlich der Interpretation von Sensoraufzeichnungen menschlicher Bewegung. Bewegungsanalyse ist nötig, um Anwendungen sensorbasierten Feedbacks in Feldern wie Rehabilitation, Sport und Ergonomie zu ermöglichen.

Bewusst kontrollierbare, zielgerichtete Bewegungen, die wir primäre Bewegungen nennen, führen aufgrund von mechanischen und physiologischen Abhängigkeiten zu kleinen, unbeabsichtigten Bewegungen in anderen Teilen des Körpers (sekundäre Bewegungen). Diese Arbeit führt Methoden zur Unterscheidung zwischen primären und sekundären Bewegungen ein. Diese Unterscheidung ist nötig, um hochqualitatives Feedback bei Aktivitäten zu liefern, bei denen ein signifikanter Anteil an sekundären Bewegungen auftritt. Dies können Aktivitäten sein bei denen große sekundäre Bewegungen aufgrund von großen Reaktionskräften auftreten (wie z. B. beim Schießen eines Balles) oder Aktivitäten bei denen die primären Bewegungen klein sind (wie z. B. bei verschiedenen handwerklichen Tätigkeiten oder beim Spiel eines Musikinstruments). Desweiteren kann eine präzise Unterscheidung zwischen primären und sekundären Bewegungen notwendig sein, um zu überprüfen, ob der Nutzer einen Teil des Körpers stillhält, wie es beispielsweise von einer gymnastischen Übung gefordert sein könnte. Von Anwendungen senorbasierten Feedbacks abgesehen ist die Verbesserung von existierenden Verfahren zur Gestenerkennung ein weiteres Anwendungsfeld für unsere Methoden. Dies ist möglich, indem die sekundären Bewegungen im Sensorsignal ignoriert werden, um zu vermeiden, dass fälschlicherweise eine Geste erkannt wird, wo eigentlich bloß sekundäre Bewegungen aufgetreten sind.

Die Effektivität unserer Methoden wird im Kontext von Armbewegungen beim Klavierspiel gezeigt, welches ein für die Analyse besonders herausforderndes Feld ist. Die Unterscheidung zwischen primärer und sekundärer Bewegung in einem Gelenk erfolgt in Abhängigkeit der Bewegungsmessung in diesem Gelenk, der Schätzungen der Reaktionskraft beim Tastenniederdruck aus MIDI-Daten und von den Bewegungen in den anderen Gelenken des Arms. Um festzustellen an welchem Arm die Reaktionskraft beim Tastenniederdruck wirkt, muss ermittelt werden, welche Hand die Note gespielt hat. Dazu führt diese Arbeit zwei Methoden ein: Die eine Methode basiert auf MIDI, die andere nutzt zusätzlich noch Daten von am Arm getragenen Inertialsensoren. Eine dritte, kamerabasierte Methode, die ursprünglich für die Erkennung von Gebärdensprache entwickelt wurde, wird in dieser Arbeit hinsichtlich der Nutzbarkeit zum Verfolgen der Hände beim Klavierspiel evaluiert.

Basierend auf unseren Analysemethoden wurden zwei klavierpädagogische Anwendungen entwickelt: Die eine Anwendung unterstützt eine existierende, klavierpädagogische Bewegungsnotation und überprüft ob die Bewegungen des Schülers der Vorgabe entsprechen. Eine Studie mit Klavierstudenten einer Musikhochschule zeigt, dass die potentiellen Nutzer das System als nützlich für das Techniktraining einschätzen. Die zweite Anwendung visualisiert Sensordaten und bietet die Möglichkeit zwei Interpretationen desselben

Stücks zu synchronisieren. Dies erleichtert es, Bewegungsunterschiede zwischen verschiedenen Spielern zu identifizieren, was Ansatzpunkte für eine genauere Untersuchung der Bewegung aufzeigt.

Acknowledgements

First, I want to thank my advisor, Max Mühlhäuser, for his guidance and for providing me with the opportunity to perform this research. I am also very grateful for my co-supervisor Paul Lukowicz, who kindly agreed to examine this work.

Furthermore, I am indebted to my college Erwin Aitenbichler for his suggestions and advice especially during the early stages of my dissertation. I thank François Lefebvre–Albaret from IRIT in Toulouse for the collaboration on applying Computer Vision methods to track pianist hands. I thank Bernhard Wetz from the music university Frankfurt for his encouragement, for our discussions, and for his help to organize the user study with piano students.

An important part of the dissertation experience was the collaboration with colleges and students on topics of mutual interest such as context architectures, in-door navigation, recognition of conducting gestures and dance patterns. For that I thank my colleges Fernando Lydaret and Erwin Aitenbichler and my students Andreas Höfer, Henning Pohl, and Diego Wong Szeto. I am very grateful for the opportunity provided by my advisor and Jan Borchers from RWTH Aachen to meet and work with the Vienna Philharmonics for which I realized a part of a new interactive conducting system. I also thank the development team from RWTH Aachen for the smooth collaboration in that project. I thank my roommates and all the colleges in the Telecooperation lab for the fun and the friendly working atmosphere.

I thank my family, my parents Birgit and Kyriakos, my brother Stephanos, and my wife Won Kyoung for their love and their continuous support during the last years and I thank my baby son Philipp for his good sense of humor.

Contents

Chapter 1.

Introduction

Recent advances in sensor technology provide new opportunities for applications that utilize the user's movement as an input source. This thesis focuses on movement analysis, which is a sub-area of a larger field concerned with the interpretation of sensor data signals of human movement, where artificial intelligence methods are often used to cope with the complexity of the signals. Movement analysis can be used to provide sensor-based feedback for training motor skills. Some major application areas are:

- **Rehabilitation:** Sensor-based feedback is expected to reduce the time to relearn motor skills, to provide a more frequent and effective training, to motivate the users, and reduce the workload for the professionals, which would also result in lower overall costs [61].

- **Sports:** Sensor-based feedback is expected to help the athletes to gain a better understanding of the movement and help them to make the appropriate corrections [167]. This is important as minimal differences can decide over success [5].

- **Ergonomics:** Performing a task adequately can help to decrease the risk of injury. A well-known example is lifting heavy objects, where advantages and disadvantages of various techniques have been discussed thoroughly in literature [44]. Sensor-based feedback can help the user to avoid injuries by warning when inexpedient movements are used.

Despite the potential benefits, sensor-based feedback is not widely used in fields mentioned above. The main technical reason for this is that current movement analysis methods are not good enough to provide sufficiently valuable feedback. Therefore, the development of new analysis methods is a valuable field for scientific study.

Goal-directed movements in one part of the body lead to slight movements in other parts of the body that are beyond direct conscious control. We call these unintended movements "secondary movements". A major contribution of this thesis are methods to distinguish between primary and secondary movement with high sensitivity. The importance of this result is two-fold:

- **Sensor-based feedback:** Secondary movement can lead to problems when providing sensor-based feedback to teach motor skills: If the secondary movement influences the feedback distinctively, the user will have the impression that the system behaves erratically. This is the case since the user cannot control secondary movement directly. By using our methods, secondary movement can be ignored by the feedback system, which solves the said problem.

- **Preprocessing step for gesture recognition:** Secondary movement could be misinterpreted by the computer to be an execution of a gesture that was not actually performed. By using our methods, it is possible to ignore secondary movement that is present in so that secondary movement cannot be mistaken as an execution of a gesture.

While methods are evaluated in context of the analysis of piano playing movements, they are generally useful to analyze a certain type of motor tasks, namely motor tasks that show a significant amount of secondary movement. To gain a better understanding for which tasks this applies, primary and secondary movements are discussed further in the following.

1.1. Primary and secondary movement

Biomechanical analysis of human movement [49] distinguishes four sources of joint torques, which are the cause for both primary and secondary movements:

- Muscular torques are torques that are generated through active muscle work [49].
- Gravitational torques are torques that are generated by the pull of gravity [49].
- Reaction torques are torques that are generated by reaction forces that are exerted on the body when interacting with the environment [49]. Consider, e. g., shooting in soccer: A counterforce acts on the foot when it hits ball. This force is then passed on to the body and generates torques in various joints, especially in the joints of the leg and the hip.
- Inter-segmental interaction torques are joint torques that are generated by movements of other parts of the body. When one limb accelerates or decelerates it exerts forces on the neighboring limbs. Movements of both distal (limbs that are further away from the center of the body) and proximal limbs (limbs that are closer to the center of the body) can produce inter-segmental interaction torques [49]. An example for a proximal inter-segmental interaction torque can be observed when setting a relaxed arm to swinging motion by twisting the upper body back and forth. An example of a distal inter-segmental interaction torque can be observed in the slight movements of the relaxed arm when wiggling the hand in the wrist up and down.

All these torques, i. e., muscular, gravitational, reaction, and inter-segmental interaction torques, can be used to execute primary movements.

Reaction torques and inter-segmental interaction torques play a central role for secondary movement: If they vary distinctively over time, they lead to displacements that cannot be controlled directly, i. e., they lead to secondary movement. The body can react on a displacement with active muscle work. Furthermore, the body can tense up the corresponding agonist and antagonist muscles pairs in preparation of a torque to reduce the amplitude of the displacement. However, if the primary movement that leads to time-varying reaction and inter-segmental interaction torques is kept, secondary movement cannot be completely avoided.

A further reason for secondary movement is that many muscles produce torques in more than one joint. The biceps brachii muscle located in the upper arm, e. g., generates movements in three joints: It supinates the forearm (it turns the palm of the hand upwards), it flexes the forearm, and is also involved (although only to a small extent) in the movements of the upper arm in the glenohumeral (shoulder) joint [23]. In that way, secondary movement can be produced in the adjacent joints when actually an isolated movement in a particular joint was intended.

Implications for sensor-based feedback: If the feedback system takes no actions to deal with secondary movement, then the degradation of the feedback quality depends largely on the proportion between primary and secondary movement: If the primary movements to execute the task are much larger than the secondary movements, feedback quality is degraded only little since the primary movement clearly dominates the sensor signal. In this case it is questionable whether an advanced handling of secondary movement is justified. However, if the secondary movements are large in comparison to the primary movements, an advanced handling of secondary movement is necessary in order to avoid a severe degradation of the feedback quality. This can be necessary for tasks with the following properties:

- Tasks where **large secondary movements** occur: Large secondary movements are usually generated when large time-varying forces act on the body. Examples are spiking in volleyball, shooting in soccer, etc. Here, a reaction force acts between the ball and the body over a short time interval. The large secondary movements produced in this way may be significant even in comparison with rather large primary movements.

- Tasks where **small primary movements** matter: Here even small secondary movements can already be significant. Examples are tasks where exact control over primary movements is necessary to achieve high precision such as various forms of handcraft and instrument performance.

- Tasks where **no primary movements** should be used in parts of the body: When a part of the body is kept still only secondary movement is present there. Instructions to keep a part of the body still can often be found in rehabilitation and gymnastic exercises.

1.2. Motivation for pianist movement analysis

The analysis of piano playing movements, more specifically the movements of the arms and the hands of the player, were the particular application field chosen to evaluate our methods. The two main applications for pianist movement analysis are:

- Piano pedagogy: Pianist movement analysis can enable new piano pedagogical applications that can provide a new path to learn technique (see Section 1.2.1).
- Electroacoustic music: Pianist movement analysis can be used to provide the pianist on stage with a way to interact with live electronics in context of compositions combining electronic and traditional acoustic instruments (see Section 1.2.2).

Pianist movement analysis is well suited to show the effectiveness of our methods as movement analysis is particularly challenging in this context:

- The primary movements are particularly small: Consider, e. g., the execution of a chord: When playing a chord, the hand is usually used as a unit [52, p. 145]. The movement to press down the keys is mainly generated by the arm and the hand. To fully depress a key, it has to be moved downwards a distance of about 1 cm. Obviously, very small arm movements are sufficient.
- The secondary movements experienced in piano playing are relatively large: When pressing down a key, an opposite force is exerted from the key to the finger. This reaction force is transmitted via the finger to the arm and can generate a considerable amount of secondary movement in the arm, which is often well visible to the human eye.

By showing that our methods performs well in context of piano playing, strong evidence is provided that they are similarly able to analyze other motor tasks that show a significant amount of secondary movement.

1.2.1. Piano pedagogy

Sensor-based feedback can provide a new way to study piano technique. Technique is only one of several areas of study addressed in piano pedagogy or, more generally, in instrument pedagogy—but an important one: Advanced students typically spend a lot of attention on technique. Evidence for this is provided by Young et al., who examined a series of instrumental lessons taught at the music department of an English university and analyzed the lesson content. Based on counting the spoken words they showed that technique is the predominant area of study: More than half of the spoken could be attributed to it [190].

Gerig characterized two complementary approaches to piano technique: the empirical approach and the analytical approach. Through practice and experimentation, the empiricist finds ways to express his musical goals while directing his concentration primarily on the musical outcome. This can be effective, which is shown by the pianists and pedagogues that have learned and taught in a primarily empirical way [54, pp. 4–6]. The analyst, who also acknowledges the importance of intensive practice, would also suggest to study analytical aspects of technique assuming that "[c]onscious, physiological introspection" [54, p. 5] during practice can be beneficial for the student's development [54, p. 5]. There has been criticism that the analytical approach is ineffective and that it can "spoil the freedom, the spontaneity, the freshness of the musical interpretation" [54, p. 4]. However, as the piano pedagogue Schultz objected, "It is one thing to say that scientific curiosity is vastly different from subjective sensibility, and quite another to imply that an extreme curiosity and an abundant sensibility cannot coexist in the same personality" [cited in 54, p. 6]. The large amount of existing works on piano technique [54, p. 5] demonstrates that many pedagogues think that analytical aspects of piano technique are important and worth studying.

Sensor-based feedback can provide an immediate, objective feedback on the student's movements. This may improve the student's ability to perform conscious introspection about the used playing movements. Thus sensor-based feedback supports the analytical side of piano pedagogy. During normal practice, the student can then apply this ability to work on technical problems he faces in the pieces he is currently studying.

While the application area of pianist movement analysis considered in this thesis is mainly piano pedagogy, pianist movement analysis can also be useful to enable new ways to interact with live electronics for the pianist on stage in context of electroacoustic music.

1.2.2. Electroacoustic music

Werner-Eppler, who was part of the group around the composer Stockhausen and the studio for electronic music in Cologne, coined the term "authentic composition" [cited in 169, p. 277], which refers to works that are independent of an actual rendering by a performer since they are stored by the composer in an authentic form on a sound carrier [169, p. 277]. This has the advantage that the composer has complete control over the piece and in essence becomes an instrument maker and interpreter as well [38]. However, this often leads to a problem of reception when electronic music is played with the audience facing an empty stage [169, p. 277].

One solution to this problem is to combine electronic sound with the sound of traditional acoustical instruments. This combination makes it necessary to coordinate the human performer and the electronic sounds. There exist several possibilities:

- The performer can synchronize herself with the electronic sound, which is fixed in time.
- An additional computer operator can control live electronics to coordinate and synchronize between the performer and the electronic sound. Alternatively, switch pedals etc. can be used by the performer on stage.
- A technique called "score following" [33, p. 4] allows synchronizing the computer to the human performance [32, 175]. Current score following algorithms are able to cope with complex scores and have been used to synchronize electronic sounds in concerts with prestigious orchestras [29].

- The sound of the instrument can be used as an input to live electronics [148]. This is somewhat similar to using an effect to modify the timbre of an instrument. However, complex mappings from acoustical to electronic sounds can be used.

- The performer on stage can be equipped with additional channels to control live electronics while basically performing naturally on his instrument. In the realm of the piano, McPherson has used continuous key position measurement, which enables the player to use extended keyboard techniques such as vibrato, slow motion of the key, and sweeps along the surface of the keyboard, to control electromagnetic excitation of the piano strings [126]. Nicolls has used accelerometers and electromyography (EMG) sensors worn on the pianist's arm to enable the player to control electronically and electromechanically generated sound [138].

Pianist movement analysis methods can help to improve the level of control that can be achieved when using movement sensors to control live electronics.

1.3. Contributions and outline

Contributions: The three main contributions of this thesis are:

1. This thesis introduces methods to distinguish between primary and secondary movement. The methods are based on a probabilistic model of human movement. Based on measurements of reaction forces and of body movements in other parts of the body, the expected amount of secondary movement is estimated. By comparing the actual measurement with that estimation, it is possible to decide whether a secondary movement has occurred. The general importance of this result is two-fold: The methods can be used (1) to enable high-quality sensor-based feedback and (2) to improve current gesture recognition techniques. The methods are applied to the analysis of pianist arm movement and are evaluated in this context. By showing that our methods perform well in context of piano playing, strong evidence is provided that they are similarly usable to analyze other, in terms of analysis complexity usually less challenging, tasks.

2. Based on our movement analysis methods, two piano pedagogical applications were developed: One application visualizes the student's movements and allows comparing performances of the same piece by different players. The second application provides auditive feedback on the student's movement. An existing piano pedagogical movement notation is used to make a connection with existing piano pedagogy practice. A user study performed with students of a music university shows that potential users think that this system is useful for learning technique.

3. The distinction between primary and secondary arm movement depends on an estimation of reaction force that acts between the key and the finger. The key reaction force is estimated from the loudness data reported by a piano with MIDI interface. In order to know where the key reaction force acts, it is necessary to

determine which hand has played the note. For that purpose, two methods are introduced and evaluated: The first method is based on MIDI alone; the second method uses inertial data from wrist-worn sensors in combination to MIDI. A third method, which was originally developed for tracking hands for sign language recognition, is evaluated for tracking pianist hands.

Outline: The thesis is structured into nine chapters including this introduction. The remaining chapters are:

- Chapter 2, "Approaches to movement analysis", discusses existing movement analysis methods and studies on piano playing movements. This chapter serves to confirm the novelty of the work on distinguishing between primary and secondary movement discussed in the Chapters 4 and 6 and provides a background on secondary movement in piano playing.
- Chapter 3, "Instrument pedagogy systems", discusses existing instrument pedagogy systems and provides the background for the Chapters 5 to 8, where the main aspects of our piano pedagogy systems are discussed. While Chapter 3 provides a survey of the field, the differences from related work for the contents of the chapters 5, 7 and 8 are discussed in the corresponding chapters.
- Chapter 4, "Movement analysis", introduces our methods to distinguish between primary and secondary movement. The methods are presented in a general, non piano-specific form, which underlines their applicability for various application fields.
- Chapter 5, "Wearable sensor system", presents the design of our wearable sensor system to capture pianist arm movements.
- Chapter 6, "Pianist movement analysis", describes the application of our movement analysis methods introduced in Chapter 4 to distinguish between primary and secondary pianist arm movements. An evaluation of our movement analysis methods is presented in this context.
- Chapter 7, "Hand tracking", presents the methods to determine which hand has played a note.
- Chapter 8, "Pedagogical applications", presents the piano pedagogical applications of the wearable sensor system, the movement analysis and the hand tracking methods.
- Chapter 9, "Summary and future work", summarizes the results and discusses possible improvements and new directions.

Parts of the results presented in this thesis have been published at various conferences [2, 73, 75, 77–80, 149] and workshops [72, 74, 76].

Chapter 2.

Approaches to movement analysis

Contents

This chapter reviews movement analysis methods and studies on pianist movement. Section 2.1 examines gesture and activity recognition and biomechanical methods to analyze movement. It is shown that these methods are inadequate to distinguish between primary and secondary movements. This confirms the novelty of our work on movement analysis presented in Chapter 4. Section 2.2 discusses studies that help to understand secondary movement in piano playing better. This puts Chapter 6, which discusses how to distinguish primary and secondary arm movements in piano playing, on a firm foundation. Section 2.3 presents studies on differences between players, e. g., between experienced and inexperienced ones in order to identify and discuss opportunities for sensor-based feedback.

2.1. Movement analysis methods

This section examines techniques in the fields of gesture and activity recognition and biomechanics to determine if they can be used to distinguish between primary and secondary movement. It is shown that traditional approaches from activity and gesture recognition and biomechanics are inadequate to solve this problem.

2.1.1. Gesture and activity recognition

Mitra & Acharya defined gesture recognition as "the process by which the *gestures* made by the user are *recognized* by the receiver" [130]. A gesture is a movement of the body made in order to communicate or to interact with the system [130]. Similarly, activity recognition can be defined as the process by which the activities made by the user, e. g., walking, standing, or commuting [92], are recognized by the computer. Gesture recognition and activity recognition are closely related fields. They employ similar machine learning methods, such as principle component analysis, hidden Markov models, and many more [130, 172].

The machine learning methods used for activity and gesture recognition can be categorized into supervised and unsupervised methods. The majority of works in activity recognition has been based on supervised methods [170, p. 15]. Supervised classification depends on a labeled data set. A sample in the data set is of the form (c, s) where c is the class label and s is a vector. The problem is to predict the class label c for a new vector s based on the information in the data set [12, p. 3]. The task to distinguish between primary and secondary movement can be viewed as a classification problem. Thus it may seem that supervised classification can be used to distinguish between primary and secondary movement. However, the attempt to collect the corresponding data sets leads to an ill-defined situation as the following thought experiment shows.

Inadequacy of supervised classification: To construct a data set, one has to decide what features to use in order to distinguish between primary and secondary movement. As primary movements are usually larger than secondary movements, it is sensible to use the movement measurement, which will be denoted as y, as one of the features. As discussed in Section 1.1, factors such as reaction forces and movements in other parts of the body cause secondary movement. Therefore, it is advisable to measure and use these factors, which will be denoted as the vector x.

First, a data set D_s that contains only samples of secondary movement is collected. To do so a subject is instructed to avoid movement in a the examined part of the body. However, based on the data set D_s it is possible to construct an almost identical data set D_p of primary movements. Let (x, y) be a sample in D_s. Now let the subject repeat the movement as closely as possible. Then the newly measured factors x' are nearly identical to the original measurements x. However, instead of keeping the examined body part absolutely still, the subject adds a slight, almost invisible primary movement ϵ. Such a minimal primary movement changes the factors x (reaction forces, movements in other parts of the body) only minimally so that $x \approx x'$. Furthermore, since only minimal primary movement is added the new movement measurement y' is almost identical to y. In essence, a sample $(x, y) \in D_s$ and a sample $(x', y') \in D_p$ have been obtained which are almost identical so that we have the situation that

$$D_s \ni (y, x) \approx (y', x') \in D_p.$$

Obviously, this is a problematic combination of data sets so that using supervised classification based on these data sets is not advisable. Two objections could be stated to refute the argument made above:

- Why should the samples in D_p contain only minimal primary movements? By instructing the subject to perform larger primary movements, the situation sketched in the equation shown above would not occur since $y \not\approx y'$. However, this immediately leads to the question how large the primary movements should be. Since primary movements are consciously controlled, it is possible to choose the size of the movements deliberately. Therefore the collected data set D_p would be too a large extent arbitrary, and, in consequence, the classification boundaries would be arbitrary as well. Our approach described in Chapter 4 circumvents this problem and uses only the data set of secondary movement D_s. A primary movement is detected if a measurement exceeds the amount of secondary movement that can be expected.

- There may be other features than x and y that may be useful to identify qualitative differences between primary and secondary movement. It is however unclear which features that would be.

2.1.2. Biomechanics

Hatze defined biomechanics as "the study of the structure and function of biological systems by means of the methods of mechanics" [83]. He distinguishes between

- Biokinematics, which is the study of biological motions without considering the causes of movement (i. e., forces)

- Biodynamics, which is the study of biological motions taking account of the involved masses and forces and

- Biostatics, which is the study of biological systems at rest [83].

Biomechanics is a broad field including various topics such as the mechanical behavior of living tissue [46], blood circulation [47], human and animal motion [45], etc. Here we are interested in the biomechanics of the human musculoskeletal system, which is a broad topic in itself, covering the anatomy and mechanical properties of tissue (bones, cartilage, ligaments, tendons, muscles, etc.), techniques to measure force, movement, and muscle contraction, as well as body modeling and analysis of human movement [139].

As discussed in Section 1.1 the causes of secondary movement are time-varying joint torques due to reaction forces and inter-segmental interaction. Based on the measurement of the movement and external forces, it is possible to determine the forces and torques that act on a joint and to break down the total force and torque into force and torque components that originate from muscle activity, gravity, reaction forces, and inter-segmental interaction [49]. There are two commonly employed procedures to tackle this problem, also known as solving the inverse dynamics: the iterative Newton-Euler formulation and the Lagrangian formulation [30, pp. 165–200].

Once the joint forces and torques have been determined, it can be desirable to estimate the internal forces and torques that act on the involved anatomical structures of the musculoskeletal system (bones, cartilage, ligaments, tendons, muscles, etc.). The problem of solving for these internal forces and torques based is known as the distribution problem. Usually there are more muscles per joint than would minimally be needed to generate movement in each direction permitted by the joint. Therefore the distribution problem is a mathematically underdetermined problem, i. e., there are more unknowns than (independent) equations. To make distribution problem a determined problem, it is possible to add additional equations based on non-trivial assumptions or to reduce the number of unknowns thus reducing the complexity of the body model. Alternatively, it is possible to use numerical optimization techniques to find a solution of the distribution problem, which minimizes the cost that is indicated by a physiologically motivated cost function. In both cases, the calculated forces may predict the actual muscle activity only to a limited degree [84, 85].

Inadequacy of biomechanical analysis: Solving the inverse dynamics and the distribution problem allows estimating the forces and torques that act on the anatomical components of the musculoskeletal system. Furthermore, it is possible to examine torques generated by muscle contraction, gravity, reaction forces, and inter-segmental interaction separately. Time-varying joint torques due to reaction forces and inter-segmental interaction are a main cause of secondary movement. However, an estimation of these torques by itself is insufficient to deduce whether the measured movement is a primary or a secondary movement.

2.1.3. Classification of our method

Our method to distinguish primary and secondary movements can be understood as a particular instance of a more general technique called outlier detection. Outlier detection is a technique that has been used for various applications, e. g., to analyze physiological data to detect diseases, intrusion detection in the area of computer security, and many more [26, p. 1]. Chandol et al. define outlier detection as "the problem of finding patterns in data that do not conform to [the] expected normal behavior" [26, p. 1]. In our case the expected normal behavior is that no primary movement was used, i. e., that the measurement contains only secondary movement. Primary movement is detected if the movement measurement is sufficiently different from the expected amount of secondary movement: A primary movement is detected if the measurement is recognized as an outlier. Out method can be understood as a outlier detection technique, which is based on a biomechanical model of human movement.

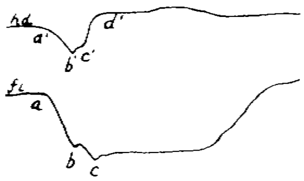


Figure 2.1.: Displacement of the hand due to key-reaction force: The upper graph indicates the movement of the hand and the lower graph, the movement of the finger. Initially, the player moves both hand and finger downwards (a–b, a'–b'). When the downward finger movement is stopped, the reaction force lifts the hand (c'–d') [141].

2.2. Secondary movement in piano playing

This section reports on empirical studies and theoretical knowledge that help to gain a better understanding of secondary movement in piano playing. Special attention is spent on the following questions:

- How is secondary movement generated in piano playing?
- How can key reaction forces be measured?
- How do these reaction forces spread over the finger, arm, and body?
- How do different playing techniques influence key reaction forces?

The first description of secondary movement in piano playing based on objective measurement that we are aware of was made by Ortmann in the 1920s and described in his book "The Physiological Mechanics of Piano Technique" [141]. Ortmann analyzed piano playing using a variety of devices that record finger and arm movement, muscle contraction, and force that is exerted to the key. While discussing action and reaction in touches, Ortmann provides the graph shown in Figure 2.1 and states:

> "When the descent of the finger-tip is suddenly arrested by the piano-key, the effect of the muscular contraction is still to decrease the angle referred to [the angle between finger and hand]. Since the fingertip cannot descend further, the only remaining way in which the angle can be reduced is by raising the hand-knuckle. The finger-tip this becomes the fulcrum and the hand-knuckle the moving part [...]" [141, p. 82]

This is what we call secondary movement.

2.2.1. Theoretical analysis

Secondary movement in piano playing is to a great extent due to the interaction with the piano action. To achieve a solid understanding of secondary movement in piano playing, a brief introduction of the piano action is necessary. This is done based on descriptions of the piano action by Wolters [182], Uchdorf [173], and Wetz [179].

The major components of the piano action: Figure 2.2 shows a schematic representation of a piano action. Conceptionally the piano action is composed of three major components: the key (1), the hammer (10), and the intermediary action, which is composed of several levers (3, 5, 9). These three major components are not mechanically fixed to each other so that they can move independently from each other. At rest, the weight of the intermediary action is supported by the key and the weight of the hammer is supported on the intermediary action.

Touch initiation: When the finger presses down the key, the rear end of the key moves upwards. The motion of the key is transmitted via the capstan (2), to the wippen (3) and leads to an upward movement of the entire intermediary action. The upward movement of the intermediary action is transferred via the jack (5) to the hammer shank (8) and makes the hammer move upwards. At a certain depth of depression, the rear end of the key touches the spoon (14), so that the damper (15) is lifted.

Escapement: If key depression is continued, the jack toe eventually comes in contact with the let-off button (4). This leads to a deflection of the jack (5), which interrupts the contact between intermediary action and the hammer. The rest of the way to the string is travelled by the hammer on its own due to its inertia. If the initial velocity of the hammer suffices, the hammer hits the string and immediately rebounds and allows the string to vibrate freely. The rebounding hammer is then caught by the back check (11). The key can be further depressed until it is stopped by the felt that is attached to the keybed.

Key release and repetition mechanism: When the player releases the key, the rear end of the key moves downwards. At a certain depression level, the back check lets the hammer free again. The repetition lever now lifts the hammer upwards using mechanical energy that was conserved in the repetition spring when the hammer rebounded. Now, the jack can glide back to its original position and reestablish the connection between key and hammer: Now a new touch can be executed. If the key is released further, the damper eventually mutes the string.

Time-varying forces: During the execution of a touch time-varying forces are exerted by the key on the finger:

- Touch initiation: As the finger begins to exert a force on the key, a reaction force acts from the key to the finger. One has to distinguish between pressed and struck touches. In a struck touch, the finger hits the key with considerable velocity from above. The impact of the finger results in a sudden maximum of key-reaction force. This impact is audible and can be used as a tone-coloring effect [52, p. 25–26]. In a pressed touch, the finger is already in contact with the key when the touch starts.

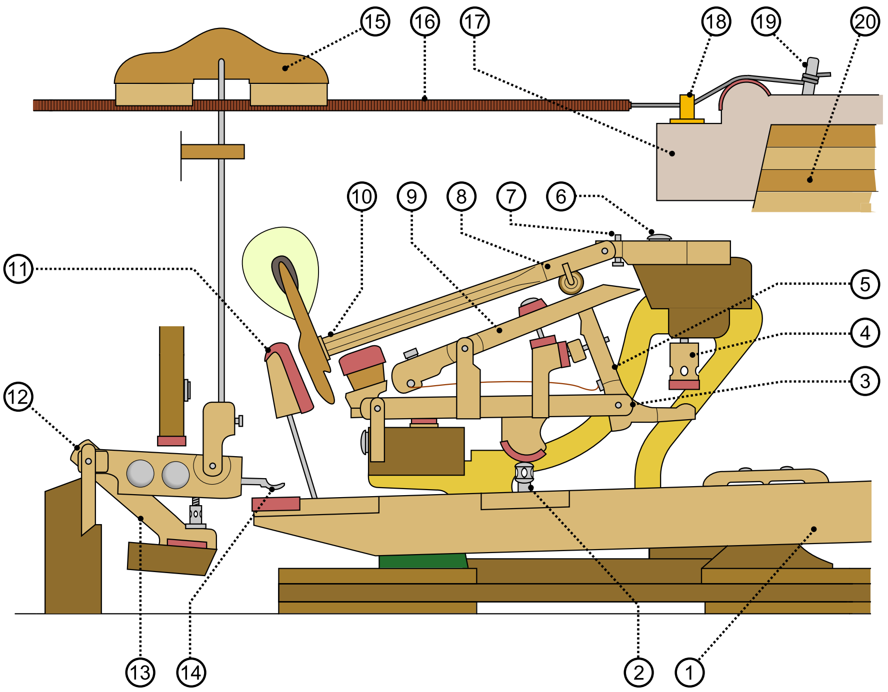


Figure 2.2.: The piano action (from `http://commons.wikimedia.org`)

- Keybed impact: When the key is fully depressed, the finger is stopped abruptly as the felt between key and keybed yields only slightly. The impact of the finger leads to a jerk with a sudden maximum in key reaction force.

- Internal impacts: Impacts inside the piano action can also have an effect on the force exerted to the finger. There is an impact when the rebounding hammer is caught by the back check. The impact is passed on by the key to the finger and changes the force that is exerted from key to finger. There are further internal impacts, namely between key and spoon and between jack toe and let-off button, that are transmitted to the finger via the key. However, these impacts are barely noticeable for the player.

- Repetition mechanism: When releasing the key, the force exerted from the key to the finger is due to the weight of the hammer, the intermediary action, and the key. At a certain depth of depression, the repetition mechanism becomes active and lifts the hammer using the mechanical energy conserved in the repetition spring. The reaction forces that are generated due to the upward movement of the hammer are transmitted via the wippen and the key to the finger.

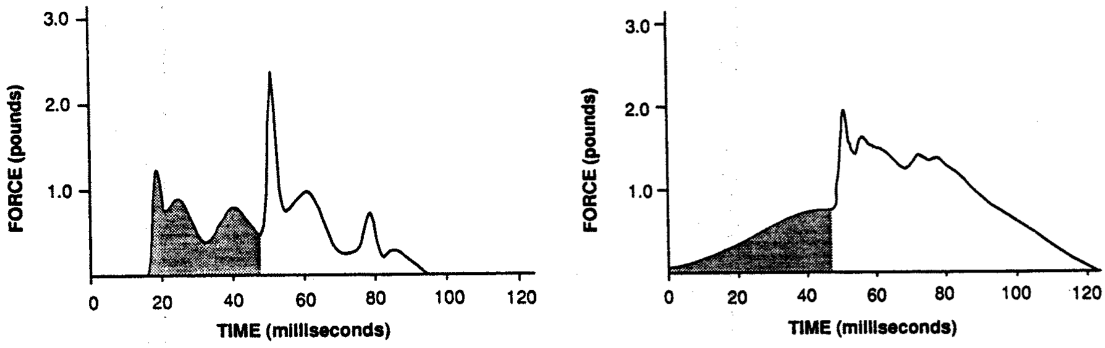


Figure 2.3.: A struck touch (left) and a pressed touch (right) [81]

2.2.2. Key reaction forces

The forces between key and finger, which are a main cause for secondary movement in piano playing have been examined by Harding et al. [81]. For this purpose, they equipped an electronic piano with MIDI interface with a piezoelectric force transducer on a white key. Several subjects with 0 to 39 years of piano playing experience executed pressed and struck touches in a variety of intensities. In a struck touch, an initial peak is visible in the force signal, which is generated by the impact of the finger on the key (see Figure 2.3, left). The highest force peak is typically generated when the key is fully depressed. The pressed touch shows a gradual increase in force (see Figure 2.3, right). The maximum peak is typically generated when the key is fully depressed. An important conclusion from Harding's et al. study is that struck and pressed touches produce different force graphs [81], which make us expect that the two touch types generate different secondary movements.

Harding's et al. results were later confirmed in a comparable experimental design by Kinoshita et al. [106]. They measured key-reaction force with a force transducer, which was installed on the key surface. Additionally, vertical key movement and sound were recorded. The pianists played a series of slow octave repetitions at different loudness levels with staccato articulation. The touches were played either with a struck or a pressed touch. Struck touches show a rapid increase of key-reaction force and a rugged key-reaction force profile while pressed touches result in a smooth key-reaction force profile. For both touch types, the key reached the lowest vertical position after the maximum of the force signal. To assess the efficiency of a touch, the impulse (which is the accumulated force over time) before reaching the lowest key position was set into proportion with the total impulse. The efficiency of a touch is highest when playing softly and becomes less efficient when playing louder. A loud note played with a struck touch shows a higher impulse than a note of same loudness played with a pressed touch [106]. Since the impulse is higher when executing struck touches, it is reasonable to assume that struck touches lead to larger secondary movement.

Parlitz et al. used pressure-sensitive sensor-matrix-foils to determine the forces exerted by the fingers on the keys [145]. The sensors-matrix-foil replaced the felt that is usually placed between the key and the keybed. Therefore they recorded only the pressure a finger exerts after the key is fully depressed. Expert pianists and novices participated in a study where they performed a series of tied-finger exercises: The player had to hold down several keys down with the inactive fingers while the active fingers executed touches. At the same tempo and loudness, the active fingers of the novices stayed in contact with the keybed for a longer time and exerted more force to it. Furthermore, the inactive fingers of the novices applied more force to the keybed [145]. Since novices apply greater forces on the keybed, it may be that they also show a greater amount of secondary movement. However, since the study concentrated only on tied-finger exercises, which is a rather artificial pattern that occurs only very rarely in actual piano literature, we cannot be certain that beginners experience a greater key reaction force in general.

An imaginative method to determine the forces between finger and key was used by Riehle et al.: They placed the entire piano on a force plate. Since the weight of the piano is constant, changes in force were due to the forces exerted by the player [154]. A disadvantage of this method is that it is not possible to determine the force exerted by each finger finger since only the total force is measured.

Discussion: Since secondary movement in piano playing depends on the key reaction force, it is sensible to measure key reaction force as basis for estimating the amount secondary movement. In the following, the techniques to measure key reaction force discussed above are discussed with respect to their usability as a part of a sensor-based feedback system.

To be usable for piano pedagogy, force sensing has to be unobtrusive. Therefore, force transducers placed on the key as used by Harding et al. [81] and Kinoshita et al. [106] in their laboratory studies are problematic. Sensor-matrix foils placed under the keys as used by Parlitz et al. [145] are also problematic as the users are usually unable to put sensor-matrix-foils under the keys. A further disadvantage of this approach is that no forces are recorded before the key reaches the key bed. Using a force plate under the piano, which was done by Wolf et al. [181] does not require the user to modify his piano and measures key-reaction force unobtrusively. However, only the total of the forces acting on all fingers can be measured with this method. It remains unclear what force acts on each individual finger when several notes are played simultaneously. Wolf et al. estimate key reaction force using regression based on MIDI data provided by the piano [181]. Since this provides an unobtrusive estimation of key-reaction force without the need to modify the piano, we use a similar method to estimate key reaction force (see Section 6.2.1). (This discussion is summarized in Table 2.1.)

2.2.3. Distinction between direct and indirect touches

As the studies reported above have shown, the key reaction forces differ between struck and pressed touches. Since secondary movement depends on the key reaction force, it would be important to determine what type of the touch the player performed. Goebl et al. have studied the use of struck and pressed touches in piano playing and have

Table 2.1.: Methods to measure or estimate key-reaction force. It is desired that no modification (Mod. = No) inside the piano action is necessary while the key-reaction forces are recorded unobtrusively (Unobtr. = Yes).

Used by	Method	Mod.	Unobtr.
Harding et al. [81]	Force transducer on the key	**No**	No
Kinoshita et al. [106]	Force transducer on the key	**No**	No
Parlitz et al. [145]	Sensor-matrix-foil under the key	Yes	**Yes**
Riehle et al. [154]	Force plate under the piano	**No**	**Yes**
Wolf et al. [181]	Estimation using MIDI data	**No**	**Yes**

developed methods to distinguish between the two: One method is to examine the vertical trajectory of the fingertip. In a struck touch, the fingertip is decelerated due to the impact on the key. The fingertip deceleration is traceable in the sensor data and can thus be used to distinguish struck from pressed touches [58]. A disadvantage of this method is that accurate finger movement measurement is necessary, which presently cannot be obtained in an unobtrusive way. Goebl et al. track passive markers that are attached to the fingertips of the player [58].

Another way to distinguish between struck and pressed touches is to examine the movement of the piano action. As already noted by Ortmann, struck and pressed touches lead to different piano action movements [140]. Ortmann developed devices to measure the movement of the key and the hammer, and the movements of the string while vibrating to show that the sound of a tone played on the piano depends only on its loudness setting pedaling and noise effects aside. To measure key movement, he used a smoked glass and an oscillating tuning fork. The tuning engraved a trace on the smoked glass, which was firmly attached to the key. Since a tuning fork oscillates with a fixed frequency, changes of the wavelength of the trace correspond to key velocity changes (see Figure 2.4). To record hammer movement, a stylus was fixed to the hammerhead, leaving a trace on a strip of smoked paper. He showed that struck and pressed touches exhibit different key movement characteristics. When playing a pressed touch, the key is continuously accelerated while when playing a struck touch, the key is abruptly accelerated in the beginning followed by a deceleration shortly after. As the finger re-engages the key, the velocity increases again [140].

Goebl et al. used accelerometers to record key and hammer movements [57]. With the accelerometer data they could not only spot events like escapement (i. e., when the contact between jack and hammer is interrupted), hammer-string contact, finger-key contact, and finger-bottom contact but they could also distinguish between struck and pressed touches. In a pressed touch, the key and hammer velocity increase gradually. In a struck touch, the key velocity shows a sudden jerk while the hammer remains still and begins to move after several milliseconds [57].

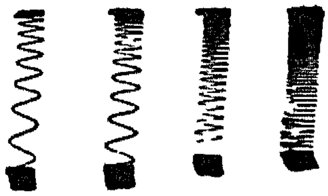


Figure 2.4.: Recordings of key movements by Ortmann. The loudness of the played tone increases from left to right [140].

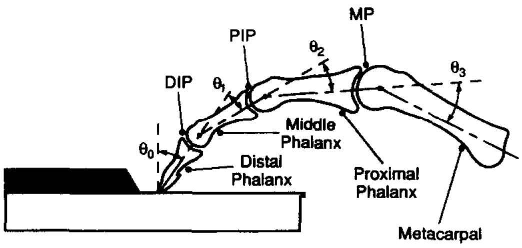


Figure 2.5.: Model of finger posture [82]

2.2.4. Distribution of key reaction forces

The key reaction forces that act on the fingertip spread out and produce forces and torques in the joints of the finger, the arm, and the entire body. These time-varying torques and forces lead to secondary movement so that it is of interest to determine them. Biomechanical methods can be used to estimate these torques and forces.

Harding et al. developed a mathematical model of the index finger that allows determining the static forces acting in the tendons and joints depending on the posture and the applied fingertip force [82]. The interaction of the muscle forces and the tendon tensions are defined based on anatomical observations. The hand posture is defined by the angles θ_i where θ_1 to θ_3 specify angles in the finger joints while θ_0 specifies the angle of the distal phalanx with respect to the key (see Figure 2.5). The angle θ_0 is used to estimate the point of contact of the key with the proximal phalanx, which changes, albeit slightly with θ_0. Assuming a quasi-static posture of the finger when the key force acts, the moments and the forces at each joint are balanced so that the key and the muscle/tendon forces can be related. An optimization technique is used to determine a posture that minimizes a joint force objective function. The minimization goal can be (1) a minimization of the linear force acting into a joint, (2) tendon tension, or (3) a combination of force and tension in the several joints and tendons [82].

Based on Harding's et al. finger model, Wolf et al. studied actual joint forces and tendon tensions in pianists while performing a repertoire piece [181]. For this purpose the performance was recorded with two video cameras. The angles θ_i were manually measured in the video and the key force was estimated from the MIDI data provided by an electronic keyboard [181].

Furuya et al. expanded the force distribution analysis by Wolf et al. (see above) to the entire arm [49, 50]. A further difference between the two works is that Wolf et al. consider only the static distribution of key reaction force that is due to the configuration of the finger while Furuya et al. consider also torques due to gravity and inter-segmental interaction (see Section 1.1). Furuya et al. distinguish the following four time-varying joint torques, which are calculated using inverse dynamics (see Section 2.1.2):

- The **gravitational torque** is generated by the weight of the limbs and depends on the posture.
- The **inter-segmental interaction torque** is generated because of the motion of the linked limbs and was calculated as a sum of centripetal, Coriolis, and inertial torques.
- The **key-reaction torque** is generated by the reactive force of the key. The key-reaction torque in a joint depends on the amplitude of the key force and the posture of the player.
- The **muscular torque** was determined as the part of the net torque that was not explained by the gravitational, key-reaction, and motion-dependent interaction torque [49, 50].

Secondary movement is generated by time-varying joint torques due to key-reaction forces and inter-segmental interaction.

2.2.5. Muscle contraction

Muscle contraction can counteract unwanted displacements and can therefore help to reduce the amount of secondary movement. On one hand, muscle force can counteract joint torques generated by reaction forces and inter-segmental interaction. On the other hand agonist and antagonist can be contracted simultaneously, which stabilizes the corresponding limb and reduces the amount of secondary movement. Some early works determined pianist muscle contraction through externally visible features like the bulging of the tendons on the back of the hand when lifting a finger [141, p. 93] or the change of the thickness of the contracting muscle [97]. Today, electromyography (EMG) based on surface electrodes is the most commonly used method to study muscle contraction in pianists.

To determine the level of contraction of a muscle with EMG, typically two electrodes that are placed above the examined muscle on the skin surface. The electrodes register potential differences that are due to electro-chemical processes in the muscle (depolarization of the muscle fibers). The potential difference between the two electrodes is amplified. The use of two electrodes has the advantage that external interference is eliminated from the signal as it occurs similarly on both electrodes. A variety of methods are used to infer information from the EMG signal. This includes the root mean square, frequency spectrum, counting of zero crossings, and counting the number of

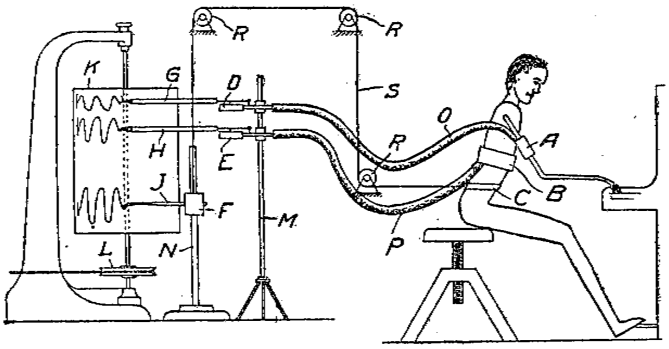


Figure 2.6.: Johnen recorded extensive measurements of pianist movements with his apparatus: Upper body movement is recorded with a belt (C). Upper arm muscle contraction and breething are recorded with pneumatic belts (A and B). The changes in air pressure are transmitted via rubber tubes to the recording device [97].

spikes above a threshold. After the EMG signal has been full-wave rectified (which can be accomplished electronically or digitally by taking the absolute of the signal), it can be averaged over a fixed window size or integrated in order to determine the level of muscle contraction. Often the resulting amplitude of the signal is normalized to the amplitude that occurs under maximal voluntary contraction [100].

Another way to assess muscle contraction is to record the change of shape of the muscles. In the 1950s Johnen used a pneumatic sleeve to record the changes in circumference in the upper arm muscles of pianists (see Figure 2.6) [97]. Today it is possible to assess muscle contraction using force sensors to detect changes in muscle shape [118].

Since secondary movement in a joint depends to a large part on the level of contraction of the corresponding muscles, works that study muscle contraction during piano performance are reviewed in the following. The studied topics were

- The influence of stress and hand posture on muscle contraction
- Differences in muscle contraction between novices and experienced players
- The effect of training on pianist's EMG signals

Influence of stress: Yoshie et al. studied the effects on physiological stress on electromyographic activity during piano performance [188]. To vary the amount of stress, the participants of the study played arpeggios in evaluation and no-evaluation settings. In the evaluation setting, the pianists could increase their cash reward by avoiding errors. Initially, each participant completed the trait sub-scale of the State-Trait Anxiety Inventory, which is an instrument to measure anxiety. Before each performance of an arpeggio, the pianists indicated their anxiety level with an "anxiety thermometer" [188], which is a continuous scale ranging from 0 (not anxious) to 100 (extremely anxious).

Heart rate was measured with an electrocardiogram (ECG) and sweat rate on the foot with a sweat rate meter based on a ventilated capsule technique. On the right forearm, the EMG signals of the extensor digitorum and the flexor carpi ulnaris were recorded and on both arms the signals of the triceps brachii and biceps brachii, and the upper trapezius. The EMG signal was full-wave rectified and low-pass filtered at 50 Hz and normalized relative to the 95 percentile over the trials in the no-evaluation mode. A measure of co-contraction was used for each agonist-antagonist muscle pair: At each sampling point in time, the minimum of the EMG signals of the agonist and the antagonist is computed. In the evaluation mode, the EMG amplitudes of all investigated muscles as well as the co-contraction measure increased [188]. Since physiological stress influences the contraction levels of the muscles, we assume that the stress level can influence secondary movement. We expect that secondary movements are slightly smaller in high stress conditions.

Differences between novices and experienced players: Bejjani et al. measured the movements of a professional pianist playing three excerpts of a piece with different hand postures [7]. EMG signals of several arm muscles, MIDI, audio, and video signals from two cameras were recorded. To identify the absolute positions in the video, markers were attached to the player's fingers and arms. The EMG signals of the following muscles of the right arm were recorded: (1) the anterior deltoid and posterior deltoid, which are shoulder muscles, (2) the triceps brachii and the biceps brachii, which extend and flex the arm in the elbow joint, (3) the flexor carpi radialis and the flexor carpi ulnaris, which simultaneously flex the hand downwards and adduct it towards the little finger, (4) the extensor carpi radialis, which simultaneously extends the hand upwards and abducts it towards the thumb, and (5) the extensor digitorum communis, which extends the four fingers of the hand. The same excerpt was played with different hand postures. Except for the deltoid and the flexor carpi radialis, no conclusive trends were evident. The deltoid activity depended on the hand posture and the played excerpt. The flexor carpi radialis was less active when the excerpts were played with stretched fingers [7].

Effects of training: Lai et al. studied differences in the EMG signal of the intrinsic muscles of the hands between pianists and non-musicians [114]. The EMG signal of the first dorsal interosseous, which is a finger muscle entirely located in the hand, was measured at different degrees of contraction. The produced force was measured with a force transducer. Automatic Decomposition EMG was used so that single motor units could be identified more clearly. At minimal contraction, the motor unit potentials of the pianists showed a higher firing rate, were shorter, and had a higher amplitude. The pianists showed higher firing rates at 25% and 50% of maximal voluntary contraction and a higher amplitude at maximal voluntary contraction than the non-musicians. Lai et al. explain these effects by neural adaptions and strengthening of the muscles through exercise in the pianists. The authors conclude that the higher firing rate of the pianists' motor units at minimal contraction indicate that they use smaller motor units, which provide a more delicate fine motor control [114].

Furuya et al. examined the differences in the EMG signal between novices and expert piano players when performing octave repetitions using forearm movement [51]. In the downswing phase, the expert players have a triceps activation that is close to resting level. The activation of the biceps muscle, the antagonist, decreases with loudness. This suggests that expert players use gravity instead of active muscle force to execute the movement. The novices however showed a distinct activation of the triceps during the downward movement of the forearm, which increased in amount and duration with loudness [51]. EMG measurements showed less co-contraction in the finger flexor and extensor muscles prior to key-bottom contact in expert players [50].

2.3. Differences between piano players

This section discusses movement differences between expert and novice players. The goal is to examine the literature to identify known differences between novice and expert players.

2.3.1. Differences between experts and novices

Furuya et al. examined experts and novices performing octave repetitions [50]. For this purpose a special camera was used to track active markers, which were placed on the knuckle joint of the little finger, the wrist, the elbow, and the shoulder. Experts performing a downswing flex their upper arm forward in the shoulder joint. Simultaneously the hand and the finger are flexed downwards. The novices, however, extended their upper arm backward. As a result, experts showed a larger angle between the finger and the key and a smaller key-reaction torque in the knuckle joint [50].

Furuya et al. studied the differences of proximal-to-distal sequencing between expert and novice piano players [48]. A proximal-to-distal sequence is characterized by the sequential timing of the movements from the most proximal segment to the most distal segment. A summation of movement effect increases the velocity at the most distal segment. The mean values of the time of peak angular velocity showed a proximal-to-distal order in the expert group. The duration of the event that a proximal segment decelerates while the distal segment accelerates was longer in the expert group. Furthermore, the peak deceleration in the proximal joint was higher in the expert group. The increased duration of simultaneous acceleration in the proximal and deceleration in the distal segment and the the increased deceleration suggest that experts use inertial forces to assist the muscular forces to move the forearm and hand thus making the movement physiologically more efficient [48].

2.3.2. Differences between different experts

Sforza et al. assessed the repeatability of finger movement in pianists of different experience [164]. For this purpose they recorded the finger movements of nine pianists performing C-major scales over the range of two octaves in different tempos. Reflective markers were placed near the fingertips and were tracked with a motion capture system.

Scales played in the same tempo were superimposed at the onset of the first note. A coefficient of superimposition was computed, which reaches 100% if the two movements are exactly identical. Statistical testing shows that the scales were played with different movements by different pianists. Furthermore, the coefficient of superimposition depends on the tempo with a significant interaction between pianist and tempo [164].

A study by Goebl & Palmer examined the role of tactile information for timing accuracy [58]. For this purpose, they recorded the finger movements of 12 pianists playing short melodies in a variety of tempos along with the MIDI data provided by an electronic keyboard. Two events are used to analyze the recorded movement data:

- Key-bottom event: The key-bottom event is defined as the point in time when the fingertip reaches the lowest point in its trajectory.

- Finger-key event: If the key is struck from above, the finger is decelerated. Finger-key events are recognized if the deceleration exceeds a fixed threshold. Since it is not imperative to strike a key from above, not all touches have a finger-key event.

The 12 pianists were divided into two groups: One group showed high percentages of touches with finger-key events in all tempos and was called the high-FK group. The other group showed relatively low percentages of touches with finger-key events in slow tempos and was called the low-FK group. The percentage of finger-key events, the amount of deceleration at a finger-key event, and the loudness of the played note increased with increasing tempo. The timing accuracy in relation to the ideal length of a note and the keystroke duration, which is defined as the time difference between finger-key event and key-bottom event, decrease with increasing tempo. Low-FK pianists showed a positive correlation with the amount of tactile feedback at the finger-key event, which is determined by the amount of deceleration, and the timing accuracy of the next note [58].

Chung et al. compared the wrist movements of pianists, who had played more than 10 years with weight-playing or traditional techniques [28]. In weight playing, the role of arm weight, which can be used to generate movement, is emphasized. Furthermore, it is emphasized in weight playing to hold the arm in place not solely with the muscles of the shoulder but also by the reaction forces that act on the fingertip when a key is depressed [86]. The two groups played exercises and samples from classical literature. Wrist movements were recorded with custom-built biaxial goniometers. The range of wrist movement in both directions was smaller in the weight-playing group. However, the energy of the movement signal was higher for the extension-flexion signal and lower for the radioulnar deviation signal when using weight-playing technique [28].

In a case study with a pianist, Wristen et al. compared the movements when performing a repertoire piece with the movements when sight-reading [183]. The angular velocities and the absolute displacements in the shoulder, elbow, wrist, and the index finger were determined based on data from a passive-marker based motion capture system. The pianist showed a greater range of angular displacement and angular velocity when performing the repertoire piece. Furthermore, the pianists used a greater range of angular displacement and angular velocity when performing the sight-reading piece a second time [183].

Ferrario et al. measured the movements of the right hand and fingers on 19 players (a concert pianists, piano teachers, and piano students) playing the first 16 measures of a minuet [41]. They used a motion capture system to track reflective markers. Five markers were placed on the nails of the fingers, one on the back of the hand, and one on the forearm near the wrist. The three-dimensional velocities of hand and fingers were used to compute the squared velocity, which is measured in units of $m^2\,s^{-2}$ and is called "unitary kinetic energy" [41] because of its resemblance to kinetic energy, which is measured in units of $kg\,m^2\,s^{-2}$. The total unitary kinetic energy was differentiated into useful and erratic unitary kinetic energy. Useful unitary kinetic energy is due to movements that are necessary for tone production while erratic unitary kinetic energy is due to extraneous movements that are not obligatory for tone production. The concert pianist showed an increased amount of total unitary kinetic energy and erratic unitary kinetic energy, and greater useful unitary kinetic energy per pressed key. The useful unitary kinetic energy per pressed key was different among the five fingers, with the thumb and index finger showing larger amounts of useful unitary kinetic energy [41].

2.3.3. Discussion

Most findings in the studies on differences between pianists are not usable as a foundation for sensor-based feedback although they help to understand pianist movement better. The reason for this is that the studies are mostly descriptive. E. g., when a player knows a piece better then he will typically use larger movements (see above), but obviously it is nonsensical to recommend to use larger movements to get to know a piece better. And of course, the authors did not imply this. There are, however, two results, that could be starting points for sensor-based feedback:

1. In contrast to novices, expert players show proximal-to-distal sequencing in the arm when playing octave repetitions [48].

2. Pianists that play use struck touches sparingly have less accurate timing [58].

As, Sesink noted, new learning media often fail to become commonly accepted if their underlying pedagogical concepts are not included in the pedagogical discourse considering what to learn, how to learn, when to learn, etc. [163]. To avoid these problems, we choose to support existing pedagogical concepts with our sensor-based feedback system, thus placing our system in the existing pedagogical discourse. Our sensor-based feedback system supports an existing notation for piano playing movements (see Section 8.3). As a side effect it is possible to rely on existing exercise material. Nevertheless these two studies could be used in the future as starting points for further sensor-based pedagogical experiments.

2.4. Summary

Methods from gesture and activity recognition were examined to determine if they can be used to distinguish between primary and secondary movement. Classification based on a data set with samples of primary and secondary movement, which is the method employed by most works on gesture and activity recognition, is inadequate to solve the said problem due to fundamental problems when collecting samples of primary movement. Likewise, biomechanical analysis of human movement does not solve the problem: While torques and forces acting on the musculoskeletal system can be computed (or estimated), biomechanical analysis does not provide a way to distinguish between primary and secondary movements. The discussion of studies on pianist movements serves two purposes namely to provide a better understanding of secondary movement in piano playing and to collect knowledge on differences between players that can be used as a basis for sensor-based feedback.

Empirical studies show that there are differences in key reaction force between struck and pressed touches. This provides evidence that the secondary movement differs according to the touch type. Since the estimation of secondary movement could be informed by knowing what type of touch is used, methods to distinguish pressed and struck touches were discussed. As secondary movement depends on the contraction of the muscles stabilizing the joint in question, studies on muscle activities of pianists were reviewed. They show influences of stress, hand posture, differences between novices and experts, and training on the EMG signals of pianists. Studies on differences between players were reviewed to identify and discuss opportunities for sensor-based feedback. While some of these studies provide interesting starting points for future sensor-based pedagogical experiments, we choose to support an existing, piano pedagogical movement notation to make the connection with existing piano pedagogical practices.

Chapter 3.

Instrument pedagogy systems

Contents

This chapter discusses existing instrument pedagogy systems with the motivation to serve as background for our sensor-based feedback system (discussed in Chapters 5 to 8) and to provide an extensive, up-to-date survey of instrument pedagogy systems. After a brief discussion of previous surveys (Section 3.1), the three main types of instrument pedagogy systems are identified and discussed in-depth, which are:

- **Augmented feedback systems**, which record a performance, analyze it, and provide visual, auditive, or tactile feedback (Section 3.2)
- **Demonstration systems**, which show how to play correctly using two- or three-dimensional visualization, augmented reality, or tactile stimulation (Section 3.3)
- **Exercise generators**, which produce exercises that are then practiced by the student (Section 3.4)

Motivation is an important aspect addressed by many instrument pedagogy systems and will be discussed in Section 3.5. A discussion at the end of this chapter (Section 3.6) provides a condensed overview over the presented systems. Comparisons with our work is deferred to later chapters, which describe the main aspects of our piano pedagogy system.

3.1. Previous surveys

In 1999, Brandao et al. published a survey of systems for music education [17]. The music education systems were grouped according to the learning goals. Systems for that teach the fundamentals of music, performance skills, music analysis skills, and composition skills were distinguished [17]. Brandao's et al. survey has a broader scope, namely music education in general, than this chapter, which reviews systems that teach performance skills.

Percival et al. distinguish instrument pedagogy systems that train a specific skill, e. g., fine pitch, and system that provide a complete training environment similar to a "virtual teacher" [147], which are intended for autodidactic learning [147]. In contrast to Brandao's et al.'s already dated survey and Percival's et al. work, which discusses only few systems, this chapter provides a comprehensive overview of the topic of instrument pedagogy systems.

3.2. Augmented feedback systems

When playing an instrument, the player receives a multitude of sensations that allow him to assess his playing, such as the sound, the tactile sensation when operating the instrument, the kinesthetic feeling of the movement, visual stimuli, etc. Augmented feedback adds additional information that supplements or augments the inherent feedback [160, p. 366]. Two types of augmented feedback are termed

- Knowledge of results (KR) and
- Knowledge of performance (KP).

KR is feedback on the outcome of an action; KP on the other hand provides feedback on the physical process to reach the goal [160, pp. 366–367]. In context of instrument pedagogy KR is feedback on the musical results while KP is feedback on the way the musical result was achieved, e. g., feedback on the movement or muscle tension. Concurrent and terminal feedback is also commonly distinguished [160, p. 366]: feedback provided after the performance has completed is called terminal feedback; feedback provided during the performance is called concurrent feedback [160, p. 366].

Augmented feedback systems for instrument pedagogy consist of three main components: sensing, analysis, and feedback. Sensing provides information about the performance to the system. Usually this information is analyzed before feedback is provided. Sensing, analysis, and feedback aspects of existing systems are discussed in-depth in the following.

3.2.1. Sensing

The music signal of an instrument performance can be captured using a microphone. Furthermore, for the piano there is the possibility to record the MIDI signal. However, the musical signal alone provides only a limited view on the performance. In particular movements, posture, and muscle tension are not obtainable from the music signal (at least not at an acceptable level of detail).

The demand to capture more aspects of the performance than only the music signal is already evident in early works. In the late 1920s, Ortmann tracked pianist movements with various mechanical devices and by photographing with extended exposure time to record traces of lights that were attached to the player [141]. A similar method was used by Hodgson in the 1930s to record the trajectory of the bow during violin playing [88]. Hodgson's cyclographs have been recently reactivated for instrumental pedagogy with modern motion capture technology [162] (see Section 3.2.3). Today, movements of instrumentalists are typically captured with

- Optical motion capture,
- Electromagnetic motion capture,
- Inertial sensing,
- Sensor-equipped instruments, or
- Electromyography (EMG).

Optical motion capture: Optical motion capture is based on markers, which are simultaneously tracked by multiple cameras (see Section 5.3.1). Optical motion capture has been used to provide movement feedback for stringed instrument pedagogy [137, 162] (see Section 3.2.3).

Electromagnetic motion capture: Peiper et al. use an electromagnetic field (EMF) sensing motion capture system (see Section 5.3.2) to record violin playing movements [146]. A sensor is attached to the instrument and another sensor to the frog of the bow. The captured values are translated from the global coordinate system into a coordinate system relative to the position and orientation of the violin [146]. Maestre et al. use an EMF motion capture system to track violin playing movements [24, 120]. A sophisticated calibration procedure is employed to increase the accuracy of the sensors and to configure the geometrical model of the violin and the bow. After calibration, a multitude of parameters can be estimated from the EMF motion capture data, namely transversal position, acceleration, distance to the bridge, and bow pressing force [24, 120].

Inertial sensing: Inertial sensors (see Section 5.3.5) can be used to capture playing movements of instrumentalists. Since inertial sensors are built into consumer devices such as mobile phones, strong commercial interests drive the development of the sensors so that better accuracy, better power efficiency, and cost reduction can be expected for the future.

Several works have recorded violin bow movement with inertial sensing. The R-Bow measures bow movement with a dual-axis accelerometer [171]. Young's system measures six degrees of freedom, i. e., three-axis linear acceleration and three-axis angular velocity, with a sensor unit on the bow [189]. A second sensor unit is placed on the body of the violin [189]. Rasamimanana et al. measure acceleration along two axes [152]. The K-Bow measures three-axis acceleration [125]. Wilmers et al. measure violin bow movement with a 6 degree of freedom sensor unit [180]. Großhauser & Hermann use a 5 degree of freedom sensor unit composed of a three-axis accelerometer and a dual-axis gyroscope to sense bowing movement [64, 65].

Sensor-equipped instruments: Sensor equipment can be integrated into traditional instruments to provide unobtrusive movement measurement. The Yamaha Disklavier [184] and the Bösendorfer CEUS [14] piano provide continuous key movement measurement and pedal movement measurement. The continuous key movement measurement provides limited information on the player's movement. While there have been attempts to include a MIDI interface in other instruments, these attempts did not lead to generally accepted solutions.

For violin instruments a variety of sensing solution have been developed and custom-built by various researchers. Systems based on inertial sensors attached to the bow and the body of the violin have already been discussed (see above). Askenfeld developed a system that measures bow position, bow velocity, bow force, and bow-bridge distance [4]. The measurement of bow position and bow-bridge distance is based on resistor wires that are integrated into the bow hair and the strings. Bow velocity is derived by differentiation of the position signal. Bow force is is measured using strain gauges that connect the bow hair to the tip and the frog of the bow [4]. Paradiso & Gershenfeld use EMF sensing to determine bow position and bow-bridge distance during violin and cello performance [144]. Two sine wave signals with different frequencies were applied to the ends of a resistive strip that spans the entire bow stick. To determine bow position and bow-bridge distance, the magnitude of the signal is measured with an antenna mounted behind the bridge. Bow pressure is measured with an urethane foam which is placed on the bow where the finger forces are applied [144]. Young and Grosshauser & Trautner use comparable setups to measure bow position and bow-bridge distance [66, 189].

Young uses strain gauges attached to the bow stick to measure vertical and lateral forces [189]. Großhauser uses matrix-based pressure sensors to measure the forces relevant to violin playing [62]. These sensors measure the pressure on a 2D grid. Pressure sensors were fixed to the bow, to measure the force exerted by each individual finger. Pressure sensors were attached to the fingertips to measure left hand finger forces. Furthermore, sensors were attached to the chin rest and shoulder rest of the violin to detect position change, cramping, and bad posture [62].

Guaus uses capacitive sensors mounted on the fingerborad of the guitar to track left hand finger movements [67].

Electromyography: LeVine & Irvine use electromyography (EMG) signals to provide left hand tension biofeedback for violinists [117]. The EMG signal of the finger muscles are measured with electrodes, which are attached to the student's left hand. When the measured tension exceeds a defined threshold, a clicking sound is played. The student practices until no tension is indicated by the sonification. The difficulty can be increased by decreasing the threshold. The system was used to train violinists, who previously reported left hand tension. Most subjects reported reduced tension levels, which persisted after the biofeedback training was discontinued [117].

Montes et al. use EMG biofeedback for the training of thumb at the piano with the thumb [132]. Their approach is based on a study by Cuvelier et al., which shows that pianists employ the muscles of the thumb, forearm, and elbow differently than non-trained subjects. They show separate phases in the EMG signal while executing a touch, which are not observable in non-trained subjects [31]. The muscle activity of the thumb muscle is measured with two electrodes, integrated over a time-interval of 0.4 s, and shown to the user on a screen. The approach was evaluated with 17 pianists, including advanced players and beginners. The players were separated in a biofeedback group and a control group. Both groups received weekly training over a duration of four weeks with a piano teacher. Subjects from the biofeedback group were additionally shown the integrated EMG signal. During training, both groups were instructed to execute a series of thumb touches with maximal muscle force and immediately relax the thumb. To record the progress of the subjects, pre- and posttest were conducted. In the pre- and posttest, which were identical for both groups, the integrated EMG signal was recorded without providing feedback to the subject. To analyze the results, the peak amplitude and the relaxation rate of the integrated EMG signal were determined. The biofeedback group achieved higher peak amplitude and relaxation rate values. Statistical testing shows that biofeedback training was able to increase the relaxation rate, which is characteristic for advanced pianists [132].

Riley uses EMG signals in combination to MIDI, audio, and video recordings to provide multimodal feedback for piano students [156]. Surface electrodes are used to measure the EMG signal of the student's muscle tension in the forearm. The processed EMG signal is recorded and presented to the user. MIDI data is recorded with a Yamaha Disklavier player piano, which is also used to replay the performance. For a detailed inspection, the MIDI data can be visualized as a piano roll. The video of the performance is recorded and automatically synchronized to the MIDI signal. A frame-by-frame navigation of the video allows close examination of the movements [156]. A scale analysis module visualizes the inter-onset-intervals, gaps between notes, and note volume [155].

Markerless motion capture: Bériault et al. developed a system that reconstructs the three-dimensional shape from the silhouettes recorded with a multi-camera setup [8]. The system is intended to be used to capture and evaluate piano performances. The movements are monitored with eight special cameras. The cameras operate at a high frame rate of 60 frames per second, are synchronized, and use global shutter exposition so that all pixels are measured simultaneously. A calibration method, which requires the user to wave a visual marker over the full working area, configures the cameras to a global coordinate system [8]. Silhouette extraction is based on the JSEG algorithm by Deng et al., which segments the image based on texture-color information and allows region tracking in a video [34]. For each camera, the extracted silhouette is projected into the three-dimensional space. The intersection of the projections defines the shape of the player. The color is determined for each volumetric pixel based on the camera images [8].

3.2.2. Analysis

Usually a feedback system has to process the audio, MIDI, video, and/or sensor data before feedback about the performance can be provided to the user. In the following approaches to analyze instrument performances are described. One can distinguish music signal analysis and movement signal analysis. Furthermore, there are works that combine different sensing modalities to enhance musical or movement signal analysis. This is called multimodal data fusion.

Music signal analysis

Automatic music transcription is the task to infer pitches, timing, rhythm, meter, articulation, etc. from the audio signal. While MIDI equipment can provide accurate pitch and timing for the piano, automatic music transcription is especially important for pedagogy systems for other instruments. A comprehensive overview can be found in a recent book edited by Klapuri & Davy [107].

Score-following is a technique to synchronize a performance to a given score [32, 175]. While the early score-following methods depended on MIDI data, it is possible to track score position based on audio signals today [150]. Score following has been used in instrument pedagogy systems to identify errors and to provide adaptive accompaniment.

Akinaga et al. developed a method that allows to evaluate scale performance of piano students [3]. Instead of comparing the performance of a scale with a mechanical performance, deviations that sound musical should not be penalized. Spline approximations for onset timing, velocity, and duration are computed for good-sounding examples. The deviations of the actual performance to the approximation are described by a set of parameters. By using multiple regression analysis or alternatively k-nearest-neighbor, the method is able to predict scores similar to the scores given by a human piano teacher [3]. An extension of the approach [134] allows changing the values of the parameters that describe the difference of the actual performance to the approximation manually. The scale then is re-synthesized based on the modified features and played to the user. By increasing the values of parameters, the student increases the deviation of the rendition from the approximation and can therefore better understand his weaknesses [134].

Oshima et al. developed a method to help piano teachers to recognize what they call the transition from the "imitation to creation stage" [143], i. e., when the student stops to imitate the teachers performance and begins to find individual ways of expression. This is achieved by calculating the difference between the performance of the student and the teacher. First, the performance is separated into a left and right hand part. The inter-onset-intervals (IOI) and loudness are determined for each quarter note. If several notes are played in the interval of a quarter note, the average velocity is computed. The IOIs and velocities are then normalized by subtracting the mean and dividing by the variance for each value. The difference between the performances is computed for IOIs and velocities separately. For this purpose, the difference performance is computed by subtracting the corresponding normalized IOIs and velocities of the two performances from each other. The difference between the performances is then determined by calculating the standard deviation of the difference performance [143].

Movement signal analysis

Video signals: Burns & Wanderley developed a system that can recognize the fingerings of a guitarist [21]. The system receives video frames from a camera, which is attached to the headstock of the guitar. The system consists of three modules: finger position recognition, recognition of strings and frets, and movement segmentation. Fingertip detection is based on the circular Hugh-transform, which detects round shapes in an image. The detection of the strings and frets is based on the linear Hugh-transform. To determine the fingering, it is furthermore necessary to detect the moment of tone production. This condition is recognized when the movement of a fingertip is below a certain threshold [21].

Kerdvibulvech & Saito developed a marker-based [102] and a markerless [103] method for guitarist left hand fingertip tracking. The marker-based method uses two cameras to track the fingertips in 3D. An ARTag is attached to the neck of the guitar to determine of the guitar relative to the the cameras. Colored markers are attached to the fingertips of the player. The color of the markers are initially learned from manually segmented training images and is then adapted automatically. This reduces the amount of manual annotation and makes the method robust against illumination changes. Particle filtering is used to track the markers in 3D. In the selection stage a particle is chosen randomly according to its probability. In the predictive stage, each chosen particle is propagated by adding Gaussian noise. In the measurement stage, the particles are projected to the 2D images of the cameras using a projection matrix, which is calculated using the information obtained by the ARTag. The probability of a particle is then changed according to the learned color model. Based on the fingertip tracking, an application was developed that shows chords and lyrics and checks if the player plays the correct chords. After the performance the application shows an overall score [102].

The markerless method by Kerdvibulvech & Saito tracks fingertip positions in two dimensions [103]. Skin color is initially learned from a small set of manually segmented images and is then automatically updated. This minimizes manual effort and makes the method robust to illumination changes. A connected component labeling algorithm is used to find and label skin-colored blobs in the image. The identification of fingertips is based on their semicircular shape. Three fingertip templates, which correspond to different fingertip orientations, are matched to the results of skin segmentation. In this way, fingertip and non-fingertip pixels are distinguished [103].

Gorodnichy & Yogeswaran developed a system that identifies the hands and fingers of a pianist in a video from a camera that is mounted above the keyboard [60]. Initially, the system detects the position of the keyboard and identifies the Middle C key. Background subtraction is used to find skin color in the keyboard area. The identified skin regions are then used to update a 2D histogram of the Cr and Cb components in a YCrCb color space. Hand tracking and identification is based on templates that describe hand location, size, and velocity assuming that only gradual changes occur between frames. By identifying crevices in the hand image, the fingers are individually identified. However, the correct finger is detected only in about 50% of the cases [60].

Sensor signals: An important part of instrument pedagogy is the training of motor skills. During the student's performance, the teacher typically does not only pay attention to the sound but also to the movements and the posture of the student. Sensor recordings of performance movements provide a signal, which can be analyzed with the computer to extract information about the used movements.

Peiper et al. developed a method that distinguishes five bowing patterns using motion capture data, namely détaché, martelé, staccato, spiccato, and legato. The patterns are distinguished by a decision tree based on geometric features (such as initial bow position) and movement features (such as velocity, acceleration, and movement continuity) [146]. Rasamimanana et al. developed a method to distinguish three bowing patterns based on accelerometer data, namely détaché, martelé, and spiccato. Minimal and maximal acceleration and velocity during a bow stroke are determined and used for classification with k-nearest-neighbor. For this purpose the velocity signal is computed by integrating the acceleration signal [152]. Young developed a method to distinguish six common bowing techniques, namely accented détaché, détaché lancé, louré, martelé, staccato, and spiccato. The classification is based on 6 degree of freedom inertial bow movement sensing and measurement of vertical and lateral bow forces. The dimensionality of the sensor data is reduced using principal component analysis. A stroke is classified in the resulting low-dimensional space using k-nearest-neighbor [189].

Based on measurements of finger pressure on the bow, Großhauser et al. were able to distinguish different bowing types and to estimate the angle in the elbow joint, which is indicative of the bow position [63]. Bow position was estimated using support vector regression based on the pressure data while using elbow angle measurements, which were obtained with a custom-built goniometer, as target values for the regression algorithm. Bowing type classification was performed using ordered means models (OMMs), which is a new machine learning method related to hidden markov models. Three different bowing styles, namely martelé, détaché, and spiccato, could be distinguished in different tempi. Furthermore correct and incorrect angle and bow pressure could be distinguished [63].

A method to distinguish German and French drum grip was developed by Bouënard et al. The method identifies characteristic local extremes of the stick trajectory in the movement signal. The grips are distinguished using k-nearest-neighbor based on the timing and the height of the extrema of the stick trajectory [15].

Multimodal data fusion

Gillet & Richard use audio and video signals for automatic transcription of drum sequences [55]. Video is used in combination to audio to detect which instrument of the drum set was played at a given instance of time. For each instrument, weighted 2D masks mark the areas where motion occurs when the instrument is played. This mask can be configured by the user or automatically learned by the system. The masks are learned by using only audio features for instrument classification and identifying motion in the video. The video features are a simple motion estimator based on the difference between consecutive frames. For each instrument, the amount of change in the area indicated by the masks is determined and accumulated over a small time interval around the note onset. A support vector machine is used for instrument classification using the described video features and a set of cepstral and spectral audio features [55].

Wang et al. use the video signals to improve the transcription of violin music for their interactive Digital Violin Tutor [177]. The video signals are used to increase the accuracy of note-onset detection, which is followed by pitch detection to finish the transcription. The system relies on a front view video to capture right hand bowing movements and a side view video to capture finger movements of the left hand. Colored markers are placed on the fingertips and the bow to simplify the task. Based on the marker tracking, the system determines fingertip and bowing velocity and bowing direction. Motion-based note onset detection functions are computed based on these features and are combined with audio-based detection functions from literature [177].

Schoonderwaldt et al. combine accelerometer signals and Computer Vision to track violin bow movements [161]. Visual markers are attached to the bow and are used to identify bow direction changes. The accelerometer values in an interval between two direction changes are integrated to yield bow velocity. Because of gravity-induced drift, the computed velocity signal can deviate from zero when bow direction changes are recognized. To eliminate this effect, the velocity discrepancy is uniformly attributed to the past interval and subtracted from the signal [161].

3.2.3. Feedback

Knowledge of results

Existing KR feedback systems have use visualization or sonification as feedback mechanisms. Two major types of visualization can be distinguished: music-score-based visualization and free visualization. In music-score-based visualization, the traditional music score is annotated with additional symbols, e. g., to indicate errors made by the student. Typically, music-score-based visualizations provide terminal feedback, i. e., they provide feedback after the performance has ended. Free visualizations are not bound to the structure of a music score and show large inter-system variability.

Music-score-based visualization: Smoliar et al. proposed a MIDI data visualization system for piano pedagogy [166]. The system visualizes tempo, articulation, and dynamics of the performance. A music-score-based visualization is used to indicate the student's articulation and dynamics in order to help the student to develop better hearing and self-assessment skills. The notes in the music score are colored according to their loudness so that uneven dynamics can be recognized by frequent color changes.

Furthermore, missed and wrong notes are marked in the score. In order to compute tempo, which is computed for each hand separately, the system compares the timing of note onsets with the nominal score. The left and right hand tempos are plotted jointly in a single graph. The desired tempo is shown in the graph for comparison. Articulation is indicated by marking the length of the recorded notes in the score. Since pedaling influences the perceived articulation, the use of the pedal is indicated in the score [166].

IMUTUS uses the audio signal of a recorder performance to recognize typical errors of beginners [43]. It detects problems in instrumental control, such as air flow or intonation problems, and general musical problems, such as wrong pitches or rhythmical inaccuracies. For analysis, the recorded audio is transformed to a symbolic representation with a audio-to-MIDI converter. The signal is then compared to the nominal score to detect deviations by the player, which are shown on a music-score-based visualization [43]. The follow-up project VEMUS adds visualizations in the score to provide feedback about the sound quality (pitch accuracy, intonation, etc.) [42].

Free visualization: Ferguson developed a visualization that combines information about the harmonic content, noisiness, and pitch accuracy in a single display [40]. Harmonic content is represented by the size of four spheres that represent the fundamental and the first three upper partials. These four spheres are lined up vertically with equal spacing when the pitch is played accurately. To indicate pitch inaccuracies, the line of spheres tilts to the left or the right to indicate a flat or sharp pitch. Noisiness is indicated by a cloud of particles, which is located in a narrow area around the line of spheres when only little noise is present and in a wide area otherwise. The length of the line of spheres is lengthened and shortened to represent changes of loudness [40].

The practice tool for pianists by Goebl & Widmer [59] generates visual feedback from MIDI input, which is immediately shown to the user (concurrent feedback). The practice tool finds and visualizes reoccurring pitch patterns. The reoccurring patterns are displayed one above the other so that timing variations can be seen. Played beats are extracted from the MIDI data and displayed along with the expected beats, which are interpolated from previous beats. A chord display shows the exact timing and the intensity of notes and allows the viewer to assess asynchronies. An acoustic piano roll, which takes into account pedal and acoustic decay, visualizes the overall performance [59].

The performance worm is a visualization that can be used to analyze expressive music performance. Tempo and loudness are indicated with a dot in a 2D space [37]. Tempo and loudness are obtained by examining the audio signal. The tempo is determined based on the output of beat tracking algorithm. Alternatively, the MIDI data of the performance can be aligned to a nominal score to compute tempo. To filter out local irregularities, tempo and loudness are smoothed over a configurable window size. The performance worm can be used to visualize a variety of expressive devices such as the rendering of phrase boundaries or large scale developments [37].

McLeod & Wyvill developed a system that visualizes musical pitch of monophonic instruments [124]. The visualization shows the frequency vs. time in a 2D diagram with reference lines that indicate discrete musical pitches. Additionally, the currently sounding note is shown in music notation. A display of the raw audio data and the amplitude of the partials is also shown. The instant visualization of the played pitch can be useful for learning fretless instruments such as the violin. Furthermore, the visualization can be useful for teaching vibrato, which is based on slightly varying pitch back and forth. The visualization of the partials can help to analyze tone quality [124].

WinSingad is a system, which provides a variety of visualization derived from the audio signal to provide feedback for singing pedagogy. Available visualizations are input waveform, fundamental frequency, various spectrogram displays. Furthermore the vocal tract area is estimated from the audio signal and visualized [90].

Gkiokas et al. developed a visualization that helps clarinetists to identify bad sounds, in particular hollow, squeaky, and unstable sounds [56]. The partials of the clarinet sound are described with Gaussian distributions learned from recordings with a professional clarinetist. Since the sound characteristics depend on pitch, the Gaussian distributions are learned individually for each discrete pitch. These distributions can be used to identify squeaky and hollow sounds, since a sound with pronounced partials is perceived as squeaky while a sound with weak partials is perceived as hollow. A feature that indicates the squeakiness and hollowness of the sound is calculated based on the Gaussian distributions. This feature is used in the visualization to control the appearance of a sphere [56].

Long notes with constant pitch and intensity are often practiced by saxophone players to learn to control air pressure. To support this type of exercise, Robine et al. developed a system that visualizes pitch and intensity [158].

Smith & Johnston developed a system to support beginning guitarists [165]. During performance, the system plays an accompaniment and shows an automatically scrolling tablature. Wrong pitches and rhythmical inaccuracies are indicated in the tablature in real-time. For offline evaluation, a correct performance of the piece and the actual performance of the student are jointly displayed in a piano roll so that pitch errors and rhythmical inaccuracies are easily identifiable [165].

Iwami & Miura developed a system that supports drummers during drum loop performance [95]. Drum loop performance is a basic practice method where the drummer keeps repeating the same rhythmical pattern. By matching the performance to a nominal score, missing notes, rhythmical deviations, and intensity deviations are determined. Missing notes are visualized conjointly with graphs of timing deviations and intensity. The average and the range of timing deviations and intensity deviations are indicated for each note [95].

Multimodal feedback: The Piano Tutor teaches basic piano playing and notation skills to beginners and provides multimodal feedback, which resemble a traditional teacher's advice, in form of video, notation, graphics, synthesized music, and voice [33]. The Piano Tutor uses score following align the incoming MIDI data with a score. This allows the system to recognize errors and inaccuracies in the student's performance. Based on an instructional design, the computer decides when the content of the next lesson based on the student's strengths and weaknesses [33].

Sonification: Ferguson's sonification studies for woodwind and brass instruments provide feedback to train fine pitch, note onset, rhythm, loudness control, legato, and vibrato [39]. When the student plays with inaccurate pitch, the sound of the closest correct note is synthesized. The sound and the amplitude of the synthesized note are modulated to indicate the magnitude of the error. A successful note onset is characterized by a short transition between silence and stable sound. The sonification indicates a successful onset with a short pentatonic melody upwards, an unsuccessful one with a melody downwards. The decision is based on user-defined time-interval. The number of notes is varied from two to five to indicate the distance from the user-defined threshold. For rhythmical training, a sound with increasing amplitude is played from the moment when an onset is expected until the note is actually played. If the player plays a note too soon, a sound is played with decreasing amplitude until the correct moment is reached. When the student plays correctly, the sonification remains silent. To train loudness control different pulsing sounds are played when the measured loudness exceeds or falls below certain boundaries. These boundaries can dynamically change over time so that continuous loudness changes can also be trained. For good legato playing woodwind players have to be attentive for gaps that can occur when depressing multiple keys. To increase the student's awareness, a sound is played when the loudness does not immediately return to the previous level. The rate of the vibrato, which is constrained between 2 and 10 Hz, is sonified by mapping it to the frequency of a partial of the currently sounding note and synthesizing this partial [39].

Knowledge of performance

Existing KP feedback systems have used visualization, sonification, and tactile stimulation as feedback mechanisms.

Visualization: The CyberViolin provides an immersive visualization of bowing motion features and bowing pattern classification data in a CAVE to allow a student to practice different articulations [146].

The Augmented Mirror (AMIR) records data from a passive marker-based optical motion capture system [137]. AMIR provides a three-dimensional visualization of the performance and allows the manipulation of the camera position, orientation, and magnification to support a close examination of the performance. Additionally, AMIR makes an audiovisual recording of the performance. The video is integrated in the visualization, where it is displayed on a two-dimensional plane that changes orientation and position according to the position of the camera [137].

A similar sensor setup is used by Schoonderwaldt & Wanderley to visualize bowing movements [162]. Similar to the cyclographs by Hodgson from the 1930s, the trajectory of the bow frog is recorded and visualized. Violin and bow are also shown in the visualization. To this end, geometrical models of the violin and the bow are used. The student can choose between two different projections so that both lateral and vertical bow movements can be seen [162].

Sonification: Rasamimanana et al. sonify string crossings to aid the violin student to execute passages with frequent string changes with even rhythm [151]. String changes are recognized with user-configurable thresholds on angular rate signals, which are measured with a gyroscope on the bow. When a threshold is exceeded, a sound is triggered. The sonification helps to concentrate the attention of the student to the string crossing task [151].

Großhauser & Hermann describe sensor-based exercises for learning the violin bow-stroke [65]. The system provides continuous auditive feedback to help the student execute the movements without unwanted deviations. The exercises consider the trajectory, the acceleration, and initial tilting of the bow. Vertical deviations from the suggested bow trajectory are expressed by modifying the frequency of a synthesized note. If the measured vertical bow trajectory deviates grossly, a second synthesized note is played to mimic the contact with a second string. Lateral deviations are expressed by changing the stereo panning of the sonification [65].

Tactile feedback: The sensor-based bowing exercises by Großhauser & Herrmann described previously were extended to provide tactile feedback [64]. For this purpose, the bow was equipped with two vibration motors, which were attached close to the position the fingers when gripping the bow. Inertial sensors and vibration motors are both integrated in the bow so that the system can be used without a computer. The feedback is generated by changing frequency and amplitude of the vibration motors. Furthermore, the two motors can be switched on and off independently [64].

3.3. Demonstration systems

Demonstration systems do not provide feedback on the student's performance but demonstrate correct playing. The majority of demonstration systems indicate how to operate the instrument and relieve the student from translating the musical score into physical actions.

Kim et al. developed a system that animates the hands of a violinist based on a musical score [105]. A best-first search algorithm is used to determine the fingering based on an evaluation function that estimates fingering effort. The trajectories of the fingertips are calculated based on the information in the musical score. The position and orientation of the wrist is determined with a neural network. The neural network was trained with the output of an optimization module, which takes into account, ease of execution, sound quality, and collision avoidance between fingers. Bow position and velocity is animated using the entire bow over a slur unless the bow velocity exceeds a predefined threshold [105].

Koutsonanos et al. developed a system that animates hand movements of a virtual recorder player [112]. The 3D animation is based on data from music notation files.

The teacher can edit and annotate the animation, which is shown to the student to demonstrate correct finger usage.

The learning assistant for electric bass uses special markers to determine position and orientation of the fingers and fretboard [22]. On a head-worn display the user can see the position of the next note on the fretboard. When the finger reaches the indicated position, the system advances the score and shows the next note [22].

PianoTouch provides tactile information to help the student to learn new pieces [91]. Five vibration motors are incorporated in a glove and provide tactile stimuli to the five fingers of the hand. When using the system the student hears the music and receives tactile stimulation of the fingers that are currently pressing the keys. Since the PianoTouch is portable it can be used for training when no piano is available [91].

Nagata et al. use motion capture and MIDI data to generate 3D animations of piano performances [113, 136]. Hand and piano key movement are animated synchronously. To capture the movements of the fingers, the pianist wears a glove with optical markers. The motion capture data is manually preprocessed to handle occlusions when fingers or hands cross. The data is then fitted to a 3D model of a hand. The MIDI data is used to animate a 3D virtual keyboard. Alternatively, Nagata et al. support the generation of synthetic movement from a monophonic musical score. The fingering is determined by minimizing a cost function using dynamic programming. The calculation of the trajectory is based on a heuristic that deals with the conflicting goals of minimizing overall movement while reaching natural playing positions. Linear interpolation is used between key frames [113, 136].

Mora et al. developed a system that overlays a 3D mesh of a suggested posture over a video of the student's performance to provide feedback on the posture [133]. The posture of a professional pianist was recorded beforehand with a visual motion capture system. The size, position, and orientation of the visualization of the pianist's body can be changed. This allows reconfiguration of the visualization so that it is well aligned with the student's body in the video. Furthermore, the student can experiment with different view angles to analyze the suggested posture [133].

3.4. Exercise generation systems

Exercise generation systems generate exercises that are subsequently practiced by the student.

Mukai et al. developed a system [135] that generates exercises similar to the exercises of Charles-Lois Hanon, a piano pedagogue of the 19th century, tailored to the weak points of the student. The system analyzes the MIDI performance of the student to identify weak points. For this purpose, the variance of the inter-onset-interval (IOI) is computed for each combination of fingers. The generated exercise consists of a four sequences of eight sixteenth notes. For each quarter note, the 3rd and 4th note of the sixteenth note subdivision contains a difficult fingering [135].

Yoo & Lee developed a system that retrieves piano exercises in response to the input of a short note sequence [186, 187]. The exercises, which are provided to the system in MIDI format, are analyzed with a repeated pattern detection algorithm to identify the typical patterns of that particular exercise. The algorithm exploits that repeated patterns show similarities in their intervallic and temporal form. The system constructs a database of exercises, which is indexed with the extracted patterns. To answer a query the system compares the query with the indices using the longest common subsequence method and delivers the most highly-ranked exercises [186, 187].

3.5. Motivation aspects

As Percival et al. noted, an important goal for many instrument pedagogy systems is to motivate the student [147]. This is sensible since the most important factor for success in motor learning is the amount of practice [160, p. 322]. Most systems address motivation as one of several goals and do so in passing: feedback provided by a system can motivate the student; likewise, the use of new technology. The Piano Tutor [33] uses a variety of media, including video and recorded speech to give the impression of a "virtual teacher" [147] to increase student motivation.

A special case is the Family Ensemble system [142], which focuses on the student's motivation. The Family Ensemble uses score following to synthesize a well-sounding accompaniment to allow parents with little musical experience to accompany their child. Playing together with the parent can motivate the child to increase practice time. A special score following-algorithm, which is robust to typical beginners' errors, analyzes the performance of the child and determines the current position in the musical score. The notes that the parent actually plays are replaced with the notes that should be heard in that particular instant of time. In consequence, the parent can press any key. Loudness and articulation are still controlled by the parent, allowing musical expression and interaction with the child [142].

3.6. Summary

The three main types of instrument pedagogy systems are:

- Feedback systems,
- Demonstration systems, and
- Exercise generation systems.

The majority of works published in scientific literature are feedback systems. Feedback systems sense, analyze, and provide feedback on the student's performance. Table 3.1 provides an overview of the feedback system that were covered in this chapter. The systems were grouped according to the musical instrument, whether the feedback provides KR or KP, the feedback modality, and whether the feedback was concurrent or terminal. In order to provide KP, other aspects of the performance beside the music signal have to be captured. Sensing options employed in existing feedback systems are optical motion capture, electromagnetic motion capture, inertial sensing, sensor-equipped instruments, and EMG. Often, the captured performance is analyzed by the computer before feedback is provided. Methods for music signal analysis and video signal analysis were discussed. The feedback modalities used in existing feedback systems are the vision, audition, and tactile modality.

Demonstration systems show how to play properly. The majority of systems show the student how to operate the instrument in order to relieve the student of the task to translate a musical score to physical actions. Table 3.2 provides an overview of the demonstration system discussed in this chapter. Vision and tactile stimulation are the output mechanisms used by current demonstration systems. The third and last category of instrument pedagogy systems are exercise systems, which generate exercises that are subsequently practiced by the student.

Table 3.1.: Overview of augmented feedback systems

System (by)	Instrument	Type	Modality	Time
Montes et al. [132]	Piano	KP	Visual	Conc.
Piano Tutor [33]	Piano	KR	Visual	Term.
Smoliar et al. [166]	Piano	KR	Visual	Term.
Oshima et al. [143]	Piano	KR	Visual	Term.
Perfromance worm [37]	Piano	KR	Visual	Conc.
Goebl et al. [59]	Piano	KR	Visual	Conc.
Riley [155, 156]	Piano	KP/KR	Visual	Term.
Akinaga et al. [3, 134]	Piano	KR	Visual	Term.
Morita et al. [134]	Piano	KR	Auditive	Term.
McLeod et al. [124]	Monophonic	KR	Visual	Conc.
IMUTUS [43]	Recorder	KR	Visual	Term.
WinSingad [90]	Voice	KR	Visual	Conc.
Ferguson [40]	Monophonic	KR	Visual	Conc.
Ferguson [39]	Wind	KR	Auditive	Conc.
Robine et al. [158]	Saxophone	KR	Visual	Conc.
Gkiokas et al. [56]	Clarinet	KR	Visual	Conc.
LeVine et al. [117]	Violin	KP	Auditive	Conc.
CyberViolin [146]	Violin	KP	Visual	Term.
AMIR [137]	Violin	KP	Visual	Term.
Cyclographs [162]	Violin	KP	Visual	Term.
Rasamimanana et al. [151]	Violin	KP	Auditive	Conc.
Großhauser et al. [65]	Violin	KP	Auditive	Conc.
Großhauser et al. [64]	Violin	KP	Tactile	Conc.
Smith et al. [165]	Guitar	KR	Visual	Conc.
Kerdvibulvech et al. [102, 103]	Guitar	KR	Visual	Term.
Iwami et al. [95]	Drums	KR	Visual	Conc.

Table 3.2.: Overview of demonstration systems

System by	Instrument	Medium
Kim et al. [105]	Violin	3D animation
Koutsonanos et al. [112]	Recorder	3D animation
Cakmakci et al. [22]	E-Bass	Augmented reality
Huang et al. [91]	Piano	Tactile stimulation
Nagata et al. [113, 136]	Piano	3D animation
Mora et al. [133]	Piano	Video, 3D body model overlay

Chapter 4.

Movement analysis

Contents

As discussed in Chapter 1, primary movements are the goal-directed movements to execute the task, while secondary movements are byproducts of the primary movements that are not under direct conscious control. Secondary movement can complicate gesture recognition and degrade the quality of sensor-based feedback for motor learning. This chapter introduces our methods to distinguish between primary and secondary movement. Based on these methods it is possible to provide high-quality feedback for motor tasks with a significant amount of secondary movement. Furthermore, the methods can be used as a preprocessing step, filtering out secondary movement from the sensor signal to improve current gesture recognition techniques.

Based on a probabilistic model of movement (Section 4.1) two methods are introduced: discrete and serial analysis. Discrete analysis (Section 4.2) is based on a data set of samples of secondary movements and allows estimating the amount of secondary movement that occurs in a fixed time-interval and to decide whether a primary or a secondary movement occurred. Serial analysis (Section 4.3) allows combining several successive discrete analyses in order to decide whether a primary movement has occurred over a larger time interval. This is possible without the need of additional data collection or training.

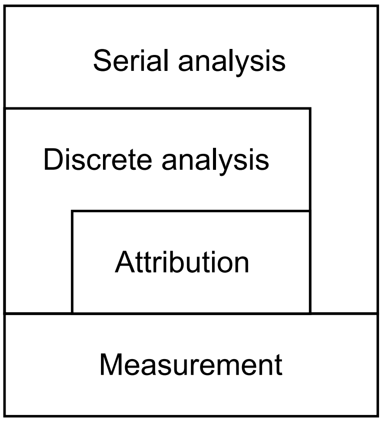


Figure 4.1.: Architecture

Architecture: The general architecture of a system using our methods is shown in Figure 4.1. The architecture consists of four components, namely measurement, attribution, discrete analysis, and serial analysis. The measurement component measures the user's movement as well as factors that influence secondary movement. Often it is sensible to measure reaction forces generated by the user on external objects as one of these factors. It may then, however, be ambiguous on which part of the body the measured force acts. Determining on which part of the body the reaction force acts is the task performed by the attribution component. Discrete analysis is based on the data provided by the measurement and attribution components. Serial analysis is based on data provided by the measurement, attribution, and discrete analysis components. This chapter discusses discrete and serial analysis in-depth.

4.1. Probabilistic Movement Model

This section introduces our Probabilistic Movement Model forms the foundation for discrete and serial analysis.

A sensor makes a measurement F of the movement in a part of the body. The measurement contains errors from several sources:

- Soft-tissue movement: The soft tissue surrounding the bones can move independently of the bone to some extent. A sensor attached to the skin experiences these motions.
- Sensor movement: The sensor can move independently of the body to some extent because of its inertia.
- Set-up inaccuracy: Inaccurate placement of the sensor can lead to inaccurate measurements.

- Sensor inaccuracy: A sensor introduces measurement errors because of its technical limitations.

Therefore, the measurement F is composed of the movement M and the measurement error E,

$$F = M + E.$$

The movement M is composed of two components: primary movement M_p and secondary movement M_s. Therefore, the measurement F can be expressed as

$$F = M_p + M_s + E. \tag{4.1}$$

Although the process that leads to secondary movement is deterministic, it is unrealistic to calculate the exact value of M_s. Therefore, secondary movement M_s is modeled as a continuous random variable. Likewise, the measurement error E is modeled as a continuous random variable.

4.2. Discrete analysis

Discrete analysis can be understood as a statistical, novelty detection method based on a biomechanical model of human movement (see Section 2.1.3). Discrete analysis estimates the probability density of $M_s + E$ in order to compare it with the actual movement measurement F. Since secondary movement depends on factors that can be measured, e. g., reaction forces, the density of $M_s + E$ is estimated in dependence of a vector x of measured factors. The density of $M_s + E$ given x is denoted $p_s(y \mid x)$ and called the conditional density of secondary movement. If the measurement F is unlikely with respect to $p_s(y \mid x)$, a primary movement is recognized; otherwise a secondary movement is recognized. In summary, discrete analysis consists of three steps:

1. **Measurement:** The movement F and the factors x are measured.
2. **Estimation:** The conditional density of secondary movement $p_s(y \mid x)$ is estimated based on measurements of the factors x.
3. **Decision:** Based on the probability of the measurement F with respect to the conditional density of secondary movement $p_s(y \mid x)$, it is decided whether a primary movement occurred.

Estimation and decision is treated in the following. Measurement, the first of the three steps, is straightforward (at least conceptually) and is not elaborated further here.

4.2.1. Estimation

Subsequently three alternative approaches to estimate the conditional density of secondary movement $p_s(y \mid x)$ are discussed: maximum likelihood estimation [12, p. 23], heteroscedastic regression [25], and quantile regression [110]. All three methods are based on a data set S with samples drawn from the density $p_s(y \mid x)$. The data set S

is collected in a controlled setting where primary movement is avoided in the examined part of the body so that a measurement F contains only secondary movement M_s and error E (see Equation 4.1). A sample $s \in S$ therefore consists of a measurement of the secondary movement and the factors of influence

$$s = (F, x) = (M_s + E, x).$$

Maximum likelihood estimation

The subsequent paragraphs describe the application of maximum likelihood estimation on the problem to estimate $p_s(y \mid x)$. Maximum likelihood is a standard statistical method and is discussed in various textbooks, e. g., in [12, p. 23].

Approach: In order to use maximum likelihood estimation to estimate the conditional density of secondary movement $p_s(y \mid x)$, the developer has to choose a parametric density type D for $M_s + E$. This could, e. g., be a normal distribution. Furthermore, the developer has to choose a parameterized function f that expresses the dependencies of secondary movement on the factors x. E. g., one may expect that secondary movements increase in size when time-varying reaction forces increase. The function f calculates the parameters of the parametric density. In the case of a normal function f would calculate the parameters of the normal distribution, i. e., mean and standard deviation. As already mentioned f itself is also parameterized. i. e., the developer provides only the general form of the function; the actual values of the parameters are determined later with maximum likelihood estimation based on the data set S.

Choosing D: A multitude of possibilities exist to choose the type of D. To determine the goodness of fit of the selected distribution D, the empirical density of $M_s + E \mid x$ with almost identical factors of influence x can be examined. For this purpose, samples of the data set in the neighborhood of an arbitrarily chosen influence factor x_0 are considered. After the empirical density has been determined, existing methods can be used to evaluate the goodness of fit, e. g., graphical methods such as histogram visualization or q-q plots [96, p. 63], or formal testing methods such as the Kolmogorov-Smirnov test [96, p. 61].

Determination of f: The parameters of the distribution D are computed by a function f based on the factors of influence x. The goal is to find a function f so that

$$p_s(y \mid x) = p(y \mid D(f(x))),$$

where $D(f(x))$ denotes the probability density that is given from the combination of the parametric density type D and the actual parameters determined by evaluating $f(x)$.

The function f is provided in a parametric form. The concrete form of f is learned from the data set S, which contains the samples $s_1, ..., s_n$ by maximum likelihood estimation. The maximum likelihood estimate of f is the concrete form of f that maximizes the density $p(S \mid f)$ of the data set S given f. The density $p(S \mid f)$ is the product of the likelihoods of the samples in the data set. The likelihood of a sample is computed by evaluating the distribution $D(f(x))$ for the sample, i. e., $p(s_i \mid D(f(s_i)))$, so that

$$p(S \mid f) = \prod_{i=1}^{n} p(s_i \mid D(f(s_i))).$$

Maximizing $p(S \mid f)$ over f yields the maximum likelihood estimate of f. To avoid underflow problems, which can occur because of the product of probability densities, the standard method of optimizing the logarithm of the likelihood is used:

$$\ln p(S \mid f) = \ln \prod_{i=1}^{n} p(s_i \mid D(f(s_i))) = \sum_{i=1}^{n} \ln p(s_i \mid D(f(s_i))).$$

Function f is a parametric function. Let w be the vector of parameters of f. Then, the maximum likelihood of f is found by maximizing over w, i. e.,

$$\underset{w}{\text{maximize}} \sum_{i=1}^{n} \ln p(s_i \mid D(f(s_i))).$$

The maximum likelihood estimate of f can be used to estimate the density of $p_s(y \mid x)$, which solves the initial problem.

Heteroscedastic regression

Another approach to estimate the conditional density of secondary movement $p_s(y \mid x)$ is to use regression. Typically, regression is used to estimate the conditional mean and assumes normal distribution with a constant variance that does not change with the independent variable x [185, p. 195]. Such regression methods are however only rarely suited to estimate the conditional density of secondary movement $p_s(y \mid x)$ as it is usually the case that not only the central moment but also the spread of secondary movement changes with the factors of influence x.

Applicability of heteroscedastic regression: Heteroscedasticity denotes the property of a conditional density $p(y \mid x)$ that the spread of the density can change with the independent variable x [185, p. 197]. In analogy, homoscedasticity is the property that the spread of a conditional density does not depend on the independent variable x [185, p. 195]. In contrast to traditional regression, heteroscedastic regression does not assume homoscedasticity and can therefore be used to analyze the conditional density of secondary movement. An extensive overview of heteroscedastic regression can be found in Carroll & Rupert's work [25] (which is the foundation for these elaborations). Although, the motivation for heteroscedastic regression is usually to provide a better estimation of the mean in heteroscedastic conditions, the conditional density $p(y \mid x)$ is often computed as a byproduct, e. g., if the conditional variance is computed along with the conditional mean (under the assumption of normal distribution). The conditional density of secondary movement can then be estimated using heteroscedastic regression with

$$p_s(y \mid x) = p(y \mid \mathcal{N}(\mu(x), \sigma^2(x))),$$

where it is assumed that $p_s(y \mid x)$ is normally distributed and where $\mu(x)$ and $\sigma^2(x)$ are the conditional mean and variance determined by heteroscedastic regression.

A widely used heteroscedastic regression method is iterative reweighted least squares [87, p. 223–224], which splits the problem of estimating the conditional mean and variance into two subproblems:

1. Given an estimator for conditional variance, the estimator for conditional mean is learnt. For this purpose, it is necessary to compute weights for each sample based on the estimated variance so that each sample gets adequate influence despite differing variance.
2. Given an estimator for conditional mean, the squared deviation ϵ^2 from the estimated mean can be computed for each sample. The conditional variance is learnt using least squares regression on ϵ^2.

The two steps are iterated until convergence is reached [87, p. 223–224].

Parametric and nonparametric regression: For the estimation of conditional mean and variance, parametric or nonparametric regression can be used. If parametric regression is used, a model has to be developed based on insight on the dependencies between the factors of influence and secondary movement, which is comparable to the formulation of the parametric function f when using maximum likelihood estimation. Nonparametric regression on the other hand learns the model from the data set, which can lead to a better fit. However, the increased flexibility also increases the danger of overfitting while parametric regression with a sound model is less vulnerable to overfitting. Furthermore, more memory is usually needed to store a nonparametric model. This can be an issue if the estimation of $p_s(y \mid x)$ has to be performed on an embedded system, e. g., as a part of a self-contained feedback system.

Comparison of methods: An advantage of using heteroscedastic regression in comparison to maximum likelihood estimation described earlier is the availability of nonparametric regression while non-parametric maximum likelihood estimation is problematic [53]. Furthermore, training is computationally less expensive. Maximum likelihood estimation is based on optimization, which can be time-intensive in training. Heteroscedastic regression on the other hand can be very efficient [115]. An advantage of maximum likelihood estimation is that it allows choosing the density type of the conditional density of secondary movement freely.

Quantile regression

Quantile regression, which was invented by Koenker & Basset [110], provides an alternative way to estimate the amount of secondary movement. An extensive overview of current quantile regression methods can be found in Koenker's recent book [109] (which is the foundation for the following elaborations). Quantile regression can be used to estimate the amount of secondary movement by estimating only the quantiles of the density instead of the entire conditional density $p_s(y \mid x)$. Koenker defines, the τ-th quantile of a density is the lowest value v for which $F(v) \geq \tau$ where F is the cumulative distribution function. With quantile regression it is possible to estimate the conditional quantiles of the conditional density of secondary movement. The τ-th conditional quantile of $M_s + E$ given x will be denoted $Q_s(\tau \mid x)$ (following Koenker's notation of conditional quantiles) and will be called the τ-th conditional quantile of secondary movement. It is the τ-th quantile of the conditional density $p_s(y \mid x)$. Parametric and non-parametric methods to estimate conditional quantiles are available [109].

Comparison of methods: In contrast to maximum likelihood estimation and heteroscedastic regression, quantile regression is not based on an assumption of an underlying probability density. In training, quantile regression is computationally less expensive than maximum likelihood estimation. However, quantile regression does not provide a full estimation of the conditional density of secondary movement $p_s(y \mid x)$ but only estimates of the conditional quantiles. This is not problematic when analyzing discrete movements. However, in order to analyze serial movements (see Section 4.3), which are composed of a sequence of discrete movements, the conditional quantiles of secondary movements are insufficient.

4.2.2. Decision

The decision whether a measurement F indicates a primary movement is based on the conditional density of secondary movement or on the conditional quantiles of secondary movement. Since a primary movements is a goal-directed movement, its size is typically larger than the secondary movement that occurs simultaneously, i. e., $|M_p| \gg |M_s|$. Based on this property, it is possible to detect primary movements.

Conditional density: To decide whether a primary movement was performed, the conditional density of secondary movement $p_s(y \mid x)$ is determined and evaluated at $y = F$. If the measurement F originates from secondary movement alone, i. e., $F = M_s + E$, then $p_s(y = F \mid x)$ yields a high value since $p_s(y \mid x)$ is the conditional density of secondary movement. However, if the measurement includes primary movement, i. e., $F = M_p + M_s + E$ and this primary movement is substantially larger than the secondary movement, i. e., $|M_p| \gg |M_s + E|$, then $p(y = F \mid x)$ yields a low value. Therefore, a primary movement is detected if

$$p_s(y = F \mid x) < c,$$

where c is a sensitivity parameter, which allows weighing between false positive and false negative errors.

Conditional quantile: To decide whether a primary movement occurred, the two conditional quantiles of secondary movement $Q_s(\tau_1 \mid x)$ and $Q_s(\tau_2 \mid x)$ where $\tau_1 < \tau_2$ are determined. A primary movement is detected if the measurement lies outside the interval spanned by the two quantiles, i. e., if

$$F \notin [Q_s(\tau_1 \mid x), Q_s(\tau_2 \mid x)].$$

By choosing the quantiles τ_1 and τ_2, the sensitivity of the detection can be controlled, which allows weighing between false positive and false negative errors.

4.2.3. Closing remarks

This concludes the introduction of discrete analysis. Discrete analysis distinguishes between primary and secondary movement. Previously this was not possible based on methods from gesture and activity recognition and biomechanical analysis (see Section 2.1). The distinction between primary and secondary movement is important to provide feedback on motor tasks that show a significant amount of secondary movement and is potentially usable to improve current gesture recognition methods (see Chapter 1).

In Chapter 6, discrete analysis is applied to analyze pianist arm movements. It is used to determine whether the player uses primary movement in the wrist, in the elbow, etc. when playing a single note. Analyzing pianist movements is especially challenging since the primary arm movements to press down a key are small while the secondary movements experienced in piano playing, which are to a great extent due to key reaction force, are large. High detection accuracy was achieved. Details are discussed in Chapter 6.

Three alternative approaches to estimate $p_s(y \mid x)$ were discussed: maximum likelihood estimation, heteroscedastic regression, and quantile regression. Later in this chapter, Section 4.3.2 discusses which approach to choose in dependency of application requirements.

4.3. Serial analysis

Temporal characteristics of discrete analysis: Each discrete analysis has two temporal parameters defining its validity in time: the length and the starting point of the discrete analysis interval. The length of the discrete analysis interval depends on the measurements contained in the data set S. As previously discussed, S contains elements of the form (F, x). The value of F is obtained on the basis of time-discrete movement sensing. Therefore, the movement measurement F spreads over a time interval that is at least as short as the time covered by a single sensor sample. Longer time intervals are also possible if several successive sensor samples are aggregated. The starting point of the analysis interval depends on the point in time when the analysis is triggered, i. e., when the factors x are measured.

Motivation for serial analysis: Serial analysis allows combining several successive discrete analyses in order to decide whether a primary movement has occurred in a larger analysis interval, which we call "the serial analysis interval". In principle, this could also be achieved with discrete analysis trained for a larger analysis interval. However, the collection of a representative data set can be prohibitively difficult when large analysis intervals are considered as the amount of possible movement variations increases with time. In consequence there is usually a ceiling where discrete analysis becomes impractical. Serial analysis evades these problems and allows detecting primary movements in analysis intervals containing several discrete analyses without the need of additional training.

Overview: Serial analysis is composed of four steps:

1. **Measurement:** The movement accumulated over the entire serial analysis interval is measured. This measurement is denoted F_{total}.
2. **Discrete analysis:** Discrete analysis is used to estimate the conditional densities of secondary movement $p_s(y \mid x_i)$ where x_i denotes the measurement of the factors influencing secondary movement for the i-th contained discrete analysis.
3. **Combination:** The estimations of conditional secondary movements $p_s(y \mid x_i)$ are used to estimate the overall density of secondary movement that occurs over the entire serial analysis interval.
4. **Decision:** Based on the probability of F_{total} with respect to the overall density of secondary movement, it is decided whether a primary movement occurred.

Measurement is conceptually straightforward. Discrete analysis has already been introduced previously. Therefore, the following elaborations concentrate on steps 3 and 4.

4.3.1. Combination

Let there be N discrete analyses contained in the serial analysis interval. We require that the beginning of the first discrete analysis interval coincides with the beginning of the serial analysis interval and the end of the last discrete analysis interval coincides with the end of the serial analysis interval. Furthermore, there may be no gaps and no overlaps of discrete analyses in the serial analysis interval, i. e., every point in time in the serial analysis interval is covered by exactly one discrete analysis. The total secondary movement $M_{s,total}$ and error E_{total} throughout the serial analysis interval is then

$$M_{s,total} + E_{total} = \sum_{i=1}^{N} M_s(i) + E(i), \tag{4.2}$$

where $M_s(i)$ and $E(i)$ denote the secondary movement and the error experienced in the i-th discrete analysis interval.

Probabilistic treatment: The variables in Equation 4.2 are treated as continuous random variables. The density of $M_s(i) + E(i)$ given the factors x_i, $p_s(y \mid x_i)$ can be determined by discrete analysis. The distribution of $M_{s,total} + E_{total}$ can therefore be computed based on algebra of random variables, which is a branch of statistics concerned with mathematical operations on random variables: The probability density of a sum s of continuous independent random variables $X_1, ..., X_N$ with probability densities $p_i(x_i)$ is given by

$$p(s) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{its} \prod_{i=j}^{N} F_t(p_j(x_j))\ dt \tag{4.3}$$

[168, pp. 57] where $F_t(f(x))$ denotes the Fourier transform of a function $f(x)$:

$$F_t(f(x)) = \int_{-\infty}^{+\infty} e^{itx} f(x)\ dx.$$

Assumption of independency: Let us consider the secondary movement and measurement error $M_s(i) + E(i)$ that occurred during two successive discrete analysis intervals $i = 1, 2$. The random variables $M_s(i) + E(i)$ are conditionally independent given the factors x_1 and x_2 if the property

$$\begin{aligned} &p\big(M_s(1) + E(1), M_s(2) + E(2) \mid x_1, x_2\big) = \\ &p\big(M_s(1) + E(1) \mid x(1), x(2)) \cdot p(M_s(2) + E(2) \mid x_1, x_2)\big). \end{aligned} \tag{4.4}$$

holds. This would imply that the value of $M_s(1) + E(1)$ does not provide additional information about the value of $M_s(2) + E(2)$ to the information provided by $x(1)$ and $x(2)$ since using the equality mentioned above yields

$$\begin{aligned} &p\big(M_s(2) + E(2) \mid M_s(1) + E(1), x_1, x_2\big) = \\ &\frac{p\big(M_s(2) + E(2), M_s(1) + E(1), x_1, x_2\big)}{p\big(M_s(1) + E(1), x_1, x_2\big)} = \\ &\frac{p\big(M_s(2) + E(2), M_s(1) + E(1) \mid x_1, x_2\big)\ p(x_1, x_2)}{p(M_s(1) + E(1) \mid x_1, x_2\big)\ p(x_1, x_2)} = \\ &\frac{p\big(M_s(2) + E(2) \mid x_1, x_2\big)\ p\big(M_s(1) + E(1) \mid x_1, x_2\big)\ p(x_1, x_2)}{p\big(M_s(1) + E(1) \mid x_1, x_2\big)\ p(x_1, x_2)} = \\ &p\big(M_s(2) + E(2) \mid x_1, x_2\big). \end{aligned}$$

Furthermore, it is usually sensible to assume that $M_s(2) + E(2)$ is conditionally independent of x_1 given x_2 since x_2 are the factors that concern $M_s(2) + E(2)$ so that

$$p\big(M_s(1) + E(1) \mid x_1, x_2) = p(M_s(1) + E(1) \mid x_1)$$

under the assumption of conditional independency from $M_s(2)+E(2)$ (Equation 4.4). In general it is not realistic to assume that $M_s(2) + E(2)$ is conditionally independent from $M_s(1)+E(1)$ given x_2 since there are usually factors that influence secondary movement that are impractical to measure in context of the aimed application and are therefore not included in x_2. Nevertheless, these non-measured factors might influence the values of $M_s(1) + E(1)$ and $M_s(2) + E(2)$ in similar ways so that the value $M_s(1) + E(1)$ may provide information about the distribution of $M_s(2) + E(2)$. Nevertheless we have to assume this independency and accept the inaccuracies that result from this simplifying assumption. Even if the conditional density $p(M_s(2) + E(2) \mid M_s(1) + E(2), x_1, x_2)$ was known, it would not be possible to use this in order to estimate the density of $M_{s,total} + E_{total}$ since the value of $M_s(1) + E(1)$ cannot be determined when analyzing the user's movement: The measurement $F(1)$ may contain primary movement as well since $F(1) = M_p(1) + M_s(1) + E(1)$ (see Equation 4.1).

Result: Under the simplifying assumption of independency (Equation 4.4) it is possible to apply Equation 4.3 to Equation 4.2 to calculate the density of $M_{s,total} + E_{total}$, which is denoted by $p_{s,total}(y)$, with

$$p_{s,total}(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ity} \prod_{i=1}^{N} F_t(p_s(z \mid x_i))\ dt, \tag{4.5}$$

where $p_s(z \mid x_i)$ is the conditional distribution of $M_s(i) + E(i)$ given x_i for the i-th discrete movement provided by the discrete analysis.

4.3.2. Decision

The measurement of the movement in the examined joint over the serial analysis interval is denoted F_{total}. The density of secondary movement $p_{total}(y)$ is computed using Equation 4.5. A primary movement is detected if

$$p_{total}(y = F_{total}) < c,$$

where c is the sensitivity parameter, which allows weighing between false positive and false negative errors.

4.3.3. Closing remarks

Serial analysis detects primary movement in larger time intervals that contain several discrete analyses. With an appropriate data set, discrete analysis can also be trained to handle larger time intervals. However, the acquisition of a representative data sets usually becomes more and more difficult as the analysis interval increases and, at some point, discrete analysis becomes impractical to use. With serial analysis it is possible to extend the analysis to larger time intervals. A disadvantage of serial analysis is that it introduces inaccuracies as it makes a simplifying assumption of conditional independency (Equation 4.4).

Serial analysis has been evaluated in the context of analyzing pianist forearm rotation movements that spread over a series of successive notes. Primary forearm rotation movements were detected with good accuracy. Details are discussed in Chapter 6.

4.4. Thresholding

Motivation: Discrete analysis allows detecting primary movements with high sensitivity. If however high sensitivity is not needed, a simple approach, where a primary movement is detected if the measurement F exceeds fixed, empirically determined thresholds, may suffice. It is worthwhile to derive this method in the framework of discrete analysis as this will provide a way to perform serial analysis based on these thresholds.

Probabilistic interpretation: In discrete analysis, the conditional density of secondary movement $p_s(y \mid x)$ is estimated. If no factors x are measured, which is the case in fixed thresholding, the (non-conditional) density of $M_s + E$ can be considered. In fixed thresholding, a primary movement is detected if the measurement F exceeds an upper t_u or lower threshold t_l, i. e., if $F < t_l \lor F > t_u$. There exists a strictly unimodal density $p_s(y)$ with the property that

$$p_s(F) < c \iff F < t_l \lor F > t_u,$$

where c is a constant. The density $p_s(y)$ should be chosen with care so that it reflects well the true density of secondary movement. The problem is similar to choosing a good density type in discrete analysis with maximum likelihood estimation (see Section 4.2.1, p. 48). The developer assess the goodness of fit of a chosen density using graphical methods such as histogram visualization or q-q plots [96, p. 63], or formal testing methods such as the Kolmogorov-Smirnov test [96, p. 61]. This derives fixed thresholding in the framework of discrete analysis, making it possible to apply serial analysis.

Result: To perform serial analysis, the density $p_{total}(y)$ of total secondary movement and error $M_{s,total} + E_{total}$ has to be determined. In analogy to Equation 4.5, the density $p_{total}(y)$ is determined with

$$p_{total}(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ity} \prod_{i=1}^{N} F_t(p_s(z))\ dt.$$

4.5. Comparison of variants

Four different variants for discrete analysis have been introduced and discussed:

- Two variants determine the conditional density of secondary movement $p_s(y \mid x)$. These are the variants based on maximum likelihood estimation and heteroscedastic regression.
- Another variant uses quantile regression and to determine the conditional quantiles of secondary movement $Q_s(\tau \mid x)$.
- Another variant uses fixed thresholds.

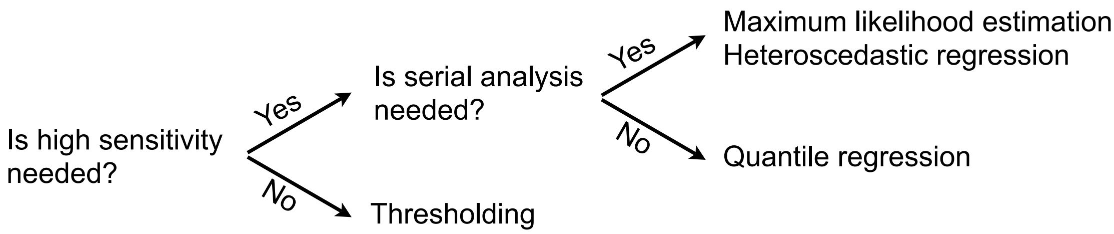


Figure 4.2.: Choosing an appropriate method

Table 4.1.: Comparison between the variants

Variant	Type of $p_s(y \mid x)$	Nonparametric
Max. likelihood estimation	Flexible	No
Heteroscedastic regression	Fixed	Yes
Quantile regression	No assumption	Yes

Two questions are important to decide what variant to use (see Figure 4.2): (1) is high sensitivity is needed and (2) is serial analysis needed? Fixed thresholding may be sufficient if high sensitivity is not needed. Otherwise one of the other variants should be used. If serial analysis is not needed, quantile regression is preferable to the other variants since it makes no assumptions on the density of secondary movement. However, if serial analysis is needed, the conditional density of secondary movement $p_s(y \mid x)$ has to be explicitly estimated so that either maximum likelihood estimation or heteroscedastic regression has to be used. An advantage of heteroscedastic regression is the availability of nonparametric heteroscedastic regression. Nonparametric maximum likelihood regression on the other hand is problematic [53]. An advantage of maximum likelihood estimation is that the density type of the conditional density of secondary movement $p_s(y \mid x)$ can be chosen freely while the type is fixed when using heteroscedastic regression (see Table 4.1).

4.6. Summary

This chapter introduced discrete and serial analysis. Discrete analysis estimates the amount of secondary movement and decides whether a primary or a secondary movement occurred. Serial analysis decodes whether a primary or a secondary movement occurred in a larger time interval that contains several successive discrete analyses.

Discrete analysis: Discrete analysis consists of three steps:

1. **Measurement:** The movement F and factors x are measured.

2. **Estimation:** Based on x, the amount of secondary movement is estimated. We introduced two ways to express the amount of secondary movement: the conditional density of secondary movement $p_s(y \mid x)$ and the conditional quantiles of secondary movement $Q_s(\tau \mid x)$. The density $p_s(y \mid x)$ can be estimated based on maximum likelihood estimation or heteroscedastic regression.

3. **Decision:** The movement measurement F is compared to the expected amount of secondary movement, estimated by $p_s(y \mid x)$ respectively $Q_s(\tau \mid x)$. A primary movement is detected if F is unlikely with respect to the estimation.

Serial analysis: Serial analysis is composed of four steps:

1. **Measurement:** The movement accumulated over the entire serial analysis interval is measured. This measurement is denoted F_{total}.

2. **Discrete analysis:** Discrete analysis is used to estimate the conditional densities of secondary movement $p_s(y \mid x_i)$ where x_i denotes the measurement of the factors influencing secondary movement for the i-th contained discrete analysis.

3. **Combination:** The estimations of conditional secondary movements $p_s(y \mid x_i)$ are combined using arithmetics of random variables to compute the overall density of secondary movement $p_{s,total}(y)$

4. **Decision:** The movement measurement F_{total} is compared to the expected amount of secondary movement, estimated by $p_{s,total}(y)$. If the probability $p_{s,total}(y = F_{total})$ is low, a primary movement is detected.

The advantage gained by serial analysis is that it allows detecting primary movements in large time intervals where discrete analysis is impractical due to problems to collect an adequate data set.

Scientific contribution: Based on our methods to separate primary and secondary movement, it is possible to provide high-quality feedback for motor tasks, which show a significant amount of secondary movement. As an exemplary case, the remaining chapters explore sensor-based feedback for piano pedagogy. However, our methods are also usable for analyzing other tasks that have a significant amount of secondary movement (see Section 1.1). Furthermore, our methods can be used as a preprocessing step to improve current gesture recognition methods, filtering out secondary movement from the sensor signal. As discussed in Chapter 2, primary and secondary movements cannot be separated with traditional supervised classification approaches commonly used in gesture and activity recognition or biomechanical analysis.

Chapter 5.

Wearable sensor system

Contents

To provide sensor-based feedback for piano pedagogy, the student's movement has to be sensed. This chapter formulates sensor system requirements (Section 5.1), examines existing sensing technologies, and describes the design of our wearable sensor system. Section 5.2 discusses why current finger movement sensing options are insufficient for application in piano pedagogy. Section 5.3 discusses options for sensing pianist arm movements. Of the available options, inertial sensing provides the best compromise between sensitivity, unobtrusiveness, and cost. However, currently available inertial sensors are either too obtrusive or too costly for the application area. Therefore, we developed our own wearable inertial sensor system called MotionNet (Section 5.4). Section 5.5 discusses practical aspects of pianist arm movement measurement with MotionNet and evaluates how well the entire measurement range of the sensing components used is exploited. MotionNet supports a variety of applications beyond piano pedagogy. Up to now, it has been used for step detection as part of an indoor navigation system [2] and for recognition of dance patterns [149].

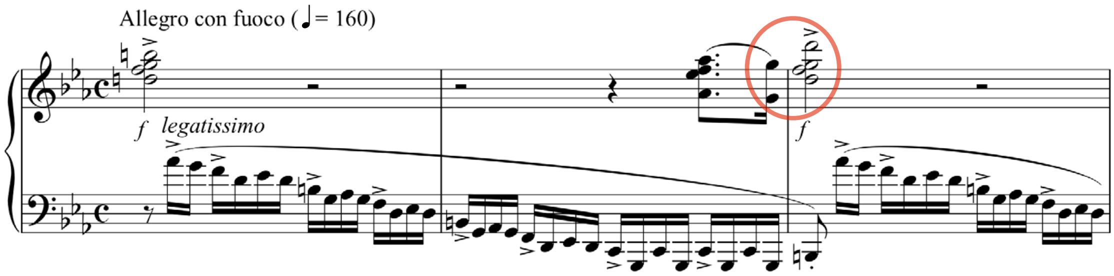


Figure 5.1.: The first bars of the Étude Op. 10 No. 12 by Chopin: When playing with the indicated tempo, the player has to play a chord, reposition the hand, and play a second chord within approximately 94 ms when playing the marked chords.

5.1. Sensor system requirements

A sensor system has to fulfill the following requirements to be usable as part of a feedback system for piano pedagogy:

- **Sensitivity (R1):** The movements that are used in piano playing are often small and quick. As an example for the small size of the movements, one may consider the execution of a chord (see Section 1.2): Only very small movements in the joints of the hand and the arm are necessary. Virtuosic piano literature contains many examples where very quick movements are necessary to produce the desired musical result. Figure 5.1 shows an example where the player has to play a chord, reposition the hand, and play a second chord within approximately 94 ms. Therefore, a sensor system has to provide high spacial and temporal resolution in order to be usable for capturing piano playing movements.
- **Unobtrusiveness (R2):** Firstly, a sensor system should not interfere with the player's movements in a way that the movement habits change or that piano playing becomes more difficult. Otherwise, the student receives feedback in context of an artificial situation, which may lead to problems when transferring the learned skills back to normal performance. Secondly, better wearing comfort can help to increase user acceptance for sensor-based feedback and is therefore also an important aspect.
- **Low cost (R3):** A large part of target users are individuals and public music schools with a limited budget. To achieve fast adoption of sensor-based feedback, low cost is important.

5.2. Finger movement measurement

Currently there are two options to capture finger movements with high accuracy: Bend-sensing gloves and passive-marker-based optical motion capture.

Bend-sensing gloves: Bend-sensing gloves [16, pp. 106–108] sense finger posture over time. Various techniques to measure finger posture are used: Light-based bend sensors rely on flexible, translucent tubes. A light emitter sends light through the tube, which is registered by a light receiver on the other side. The amount of light that reaches the receiver changes in dependency of the finger posture. A similar technique is based on a translucent tube with a reflective interior wall. A special light receiver determines the intensity of the direct light and the reflected light separately [16, pp. 106–108]. An example for a bend-sensing glove is the CyberGlove II by Immersion [93]. It provides 22 simultaneous measurements of finger and wrist posture at a rate of 100 Hz. A severe disadvantage of bend-sensing gloves is that they impede the movability of the fingers by introducing friction that has to be compensated by additional muscle force. This complicates piano playing (violation of R2). Therefore, bend-sensing gloves are not suited as basis for sensor-based feedback for piano pedagogy.

Optical motion capture: Optical motion capture systems track movements three-dimensionally based on markers placed on the body of the player [127]. Two major variants exist: systems with active markers and systems with passive markers. Active systems track infrared LED lights; passive systems flood the tracking area with infrared light and track reflective markers. In both cases the position of a marker is detected simultaneously on multiple cameras so that the 3D position can be triangulated [127].

Optical motion capture systems can provide high spacial and temporal resolution (R1). E. g., the T160 camera[1] used in the motion capture system VICON, provides a resolution of 16 megapixels at a frame-rate of 120 fps while frame-rates up to 2000 fps are possible at reduced resolution. Marker occlusion can result in measurement glitches, which can be problematic since time-intensive, manual post-processing of the data usually done in laboratory settings is not sensible in context of piano pedagogy. Misidentification of markers can be a problem when using passive systems and an additional source of measurement glitches. Active systems, such as the Optotrack by NDI,[2] can unambiguously identify markers so that marker misidentification does not occur. In contrast to passive markers active markers are uncomfortable to wear on the fingers (R2) as they need either to be connected with cables or to be equipped with a battery. Optical motion capture systems are very expensive, which would prohibit fast adoption for piano pedagogy (violation of R3).

5.3. Sensing options

As previously discussed, current sensor options are insufficient to track pianist finger movements. Arm movements are also an important concern in piano pedagogy (see Sections 8.1 and 8.3.1). In the following, options to track pianist arm movements are discussed, namely

- Optical motion capture,

[1] `http://www.vicon.com/products/t160.html`
[2] `http://www.ndigital.com/lifesciences/certus-motioncapturesystem.php`

- Electromagnetic motion capture,
- Electromechanical motion capture,
- Computer Vision based on consumer cameras, and
- Inertial sensing.

5.3.1. Optical motion capture

Optical motion capture has already been discussed previously in context of finger movement measurement. Optical motion capture provides high spacial and temporal resolution. Both active and passive markers can be worn comfortable on the arm. However, optical motion capture systems are very expensive, which would prohibit fast adoption for piano pedagogy (violation of R3).

Markerless motion capture is an active field of research as shown by the recent survey by Moeslund et al. [131]. However, in its current state markerless motion capture is not yet mature enough to be usable for piano pedagogy.

5.3.2. Electromagnetic motion capture

Electromagnetic motion capture is based on a transmitter, which generates an electromagnetic field, and several receivers worn on the body of the performer [127, p. 20]. Orthogonal coils in the receivers measure the magnetic field. The data obtained by the receivers is used to calculate the 3D position and orientation [127, p. 20].

In general, electromagnetic motion capture can provide measurements with high spacial and temporal resolution. E. g., the Liberty[3] by Polhemus provides position accuracy of 0.03 in and orientation accuracy of 0.15° root mean square at a sample-rate of 240 Hz. In contrast to optical motion capture, electromagnetic motion capture does not suffer from occlusion effects. However, because of their sensitivity to metals, which are contained in the iron-cast plate, the strings, etc., it would be highly problematic to use electromagnetic motion capture for capturing pianist movements (violation of R1).

5.3.3. Electromechanical motion capture

Electromechanical motion capture [127, p. 23] is based on a mechanical exoskeleton that is worn by the performer. Potentiometers are used to measure the joint angles. Electromechanical motion capture provides highly accurate measurements with high temporal resolution. However, it is not usable for capturing piano playing since the mechanical resistance would interfere with the student's playing (violation of R2).

[3] `http://www.polhemus.com/?page=Motion_Liberty`

5.3.4. Computer Vision

Instead of relying on expensive high-performance video cameras used in optical motion capture systems, it is possible to analyze the video signal from consumer cameras. However, the spacial and temporal resolution of consumer cameras is limited so that only large scale movements can be tracked (violation of R1).

5.3.5. Inertial and magnetic field sensors

An inertial sensor can determine its proper motion without external reference by measuring the effects of its motion on an inertial proof mass. The two main types of inertial sensors are accelerometers and gyroscopes. Accelerometers [128] measure linear acceleration. The movement of the accelerometer proof mass can be sensed with various techniques, e. g., by capacitive measurement, inductive measurement, or by measuring the deformation of a piezoelectric element [128].

Gyroscopes [129] measure angular velocity. Three main designs for gyroscopes are distinguished: rotary gyroscopes, vibrating gyroscopes, and optical gyroscopes. The mechanical rotary gyroscopes are based on the conservation of the angular momentum of a spinning body. Vibrating gyroscopes are based on the Coriolis effect. A proof mass moves back and forth on a track orthogonal to the sensing axis. When the gyroscope is rotated, the vibrating proof mass drifts sideways along the track because of Coriolis force. Optical gyroscopes are based on the Sagnac effect. A laser light beam is separated with a semi-transparent mirror into two beams. The two light beams are reflected and travel with different paths to a photodetector. When the optical gyroscope is being rotated, the distance of one path is minimally reduced while the distance of the other path minimally lengthened for the passing light. This can be sensed by measuring the interference between the two beams with the photodetector [129]. Microelectromechanical systems (MEMS) versions of accelerometers and gyroscopes, which are small and inexpensive, are available on the market.

Similarly to inertial sensors, magnetic field sensors [157] can provide information about the movement without external reference. This is accomplished by measuring the Earth's magnetic field. Anisotropic magnetoresistance (AMR) is the property of a material to change its resistance according to the direction of a magnetic field. AMR sensors are sufficiently sensitive for Earth magnetic field sensing. Giant magnetoresistance (GMR) sensors are based on the same principle. However, the layered structure of magnetic and non-magnetic layers exhibit greater anisotropic magnetoresistance effects [157].

The 3D orientation can be calculated by combining accelerometer and magnetic field sensing signals based on the direction of gravity and magnetic north [101]. Acceleration that occurs because of human movement however disturbs the accelerometer measurement of gravity. To diminish this problem, gyroscope signals can be used to calculate the short-term change of orientation. For this purpose, Kalman filtering can be used for sensor fusion [159].

Inertial sensors can provide high measurement resolution at high sample rates (R1). Our sensor provides a resolution of 9.8 mg (see Table 5.4) in acceleration measurement

Table 5.1.: Sensor properties for capturing pianist arm movements

	Sensitive	Unobtrusive	Low cost
Optical motion capture	+	+	−
Electromag. motion capture	−	∘	−
Electromech. motion capture	+	−	−
Computer Vision	−	+	+
Inertial & magnetic field sensors	+	∘	+

and 1.6 °/s in angular rate measurement (see Tables 5.5 and 5.6). The sensor chips are available in small packages [35, 36, 94] so that small inertial measurement units (IMU), which can be worn acceptably comfortable on the player's arm (R2), can be built. As the chips and the other electronic components are cheap, inertial measurement units can be manufactured at low cost (R3).

5.3.6. Discussion

Sensing technologies: Table 5.1 summarizes the advantages and disadvantages of the options for capturing pianist arm movement. Electromagnetic and electromechanical motion capture is principally not suited: Electromagnetic motion capture does not perform well in environments with metals (iron plate, strings); electromechanical motion capture is too obtrusive. Furthermore, Computer Vision with consumer cameras provides too little spacial and temporal resolution. Optical motion capture systems are less obtrusive than inertial sensors. Furthermore, they can provide absolute position information, which cannot be provided by inertial sensors. However, optical motion capture systems are expensive, which would prohibit fast adoption for piano pedagogy. Inertial sensing can provide high sensitivity (R1), can be acceptably unobtrusive (R2), and can be manufactured at low cost (R3). Therefore, we chose inertial sensing as the enabling technology for our feedback system.

Commercially available IMUs: Having decided for a technology, commercially available inertial sensing solutions were surveyed. Some systems such as the the InertiaCube,[4] use wired communication without a bus. However, to record the movements in all arm joints several sensor units per arm are needed. Without a communication bus, a separate cable connection between each sensor unit and the computer. This, however, raises the question how to connect many sensors while maintaining wearing comfort (R2). Commercially available wireless sensor platforms, like, the Crossbow,[5] have too little bandwidth to transmit sensor data with adequate temporal resolution (violation

[4] `http://www.intersense.com`
[5] `http://www.xbow.com`

Table 5.2.: MTx and MotionNet

	MTx	MotionNet
Max. number of sensors at 100 Hz	10	80
Weight	30 g	10 g
Magnetic field sensor	Yes	No
Power consumption	360 mW	170 mW
Cost (approx.)	EUR 1,750.–	EUR 150.–

of R1). The closest to our system is the XSens MTx Xbus.[6]. The XSens MTx Xbus is based on bus communication with sufficient bandwidth for high data rates (R1). Being based on a communication bus reduces the amount of necessary cabling, which results in better wearing comfort (R2). However, the XSens MTx Xbus is very expensive, which would prohibit fast adoption for piano pedagogy (violation of R3). A single MTx sensor costs approximately EUR 1,750 while a single MotionNet sensor costs approximately EUR 150. The total cost of the sensing system with three sensors per arm would total to EUR 10,500 for the MTx and EUR 900 for the MotionNet. Table 5.2 provides a comparison between MTx and MotionNet.

5.4. MotionNet

MotionNet uses bus communication to reduce the amount of cabling and increase wearing comfort. Controller Area Network (CAN), which is a serial, message-based broadcast bus, provides sufficient bandwidth of up to 1 Mbit/s to support high sensor sampling rates. CAN support is included in inexpensive microcontrollers, such as the Atmel AT90CAN128 on which MotionNet is based, together with other features needed for the realization of the sensors such as analog-digital-conversion (ADC) and processing capability. This helps to reduce the overall size of the sensor units.

Since ordinary computers do not support CAN, mediation between the sensor units and the computer is necessary. This task is performed by the host unit, which uses RS232 or Wi-Fi to communicate with the computer. Each sensor unit is equipped with a inbound and outbound CAN connector so that several sensor units can be connected in series. In order to enable different physical configurations, splitters, which have a single inbound and two outbound ports, can be used. Figure 5.2 shows three example configurations.

[6] `http://www.xsens.com`

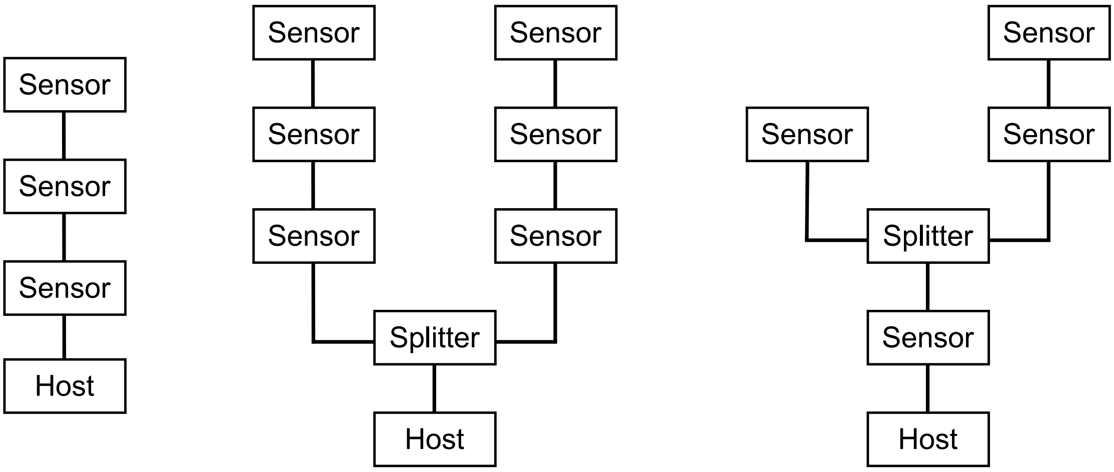


Figure 5.2.: Configuration to capture one arm using three sensors (left), configuration for two arms (center), and the configuration used for dance pattern recognition (right)

5.4.1. Sensor units

Two aspects were considered especially important when choosing the sensing elements for the MotionNet sensor units. Firstly the sensing elements should provide a good compromise between the conflicting goals to have a large measurement range and high sensitivity: The sensitivity should be sufficient so that delicate playing movements can be recorded accurately. Likewise, the measurement range should be sufficient to record quick playing movements accurately. Secondly, the sensing elements should be highly integrated to save space and increase wearing comfort. Therefore, accelerometers and gyroscopes that sense motion in several axes were preferred.

The following sensing elements are used in the MotionNet sensor units:

- The ADXL330 3-axis accelerometer from Analog Devices [36],
- The IDG-300 2-axis[7] gyroscope from InvenSense [94], used to measure pitch and roll rotation, and
- The ADXRS300 single-axis gyroscope from Analog Devices [35], used to measure yaw rotation.

Due to the space considerations, the sensor units were not be equipped with a magnetic field sensors. A sensor unit is shown in Figure 5.3. Section 5.5 shows that the measurement range is adequate for recording piano playing movements.

[7]When MotionNet was developed 3-axis gyroscopes were not available. In the meantime, this has changed.

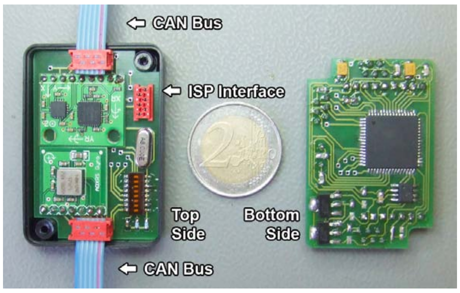


Figure 5.3.: MotionNet sensor unit

Table 5.3.: ADC properties

	Single-ended	Differential	Unit
Resolution	10	8	Bit
Absolute accuracy	1.5	1	LSB
Integral non-linearity	0.5	0.5	LSB
Differential non-linearity	0.3	n/a	LSB
Gain error (max.)	±2	±2	LSB
Offset error (max.)	±2	±1	LSB

Analog digital conversion

Analog digital conversion is performed with the analog digital converter (ADC) of the Atmel AT90CAN128 microprocessor. The ADC provides single-ended conversion, which is used to sense the accelerometer signals, and differential conversion, which is necessary to sense the gyroscope signals. Table 5.4 provides the technical characteristics of the ADC as specified by the manufacturer taking into account the specific circuitry of the sensor unit.

Accelerometer

The ADXL330 from Analog Devices [36] is a MEMS 3-axis accelerometer based on capacitive sensing of an inertial proof mass, which is suspended with polysilicon springs. Table 5.4 provides the technical characteristics of the ADXL330 as specified by the manufacturer taking into account the specific circuitry of the sensor unit. Sensitivity is expressed in least significant bits (LSB) based on the characteristics of the ADC.

Gyroscopes

The gyroscopes used on the MotionNet sensor are the IDG-300 from InvenSense, which measures pitch and roll rotation, and the ADXRS300 from Analog Devices, which mea-

Table 5.4.: Accelerometer ADXL330 characteristics

Parameter	Value	Unit
Measurement range	±3.6	g
Sensitivity	9.8	mg/LSB
Bandwidth	50	Hz
Nonlinearity	0.3	% of full scale
Noise rms of x- and y-axis	2.5	mg
Noise rms of z-axis	3.1	mg

Table 5.5.: IDG-300 2-axis gyroscope characteristics

Parameter	Value	Unit
Measurement range	±500	°/s
Sensitivity	1.6	°/s/LSB
Nonlinearity	<1	% of full scale
Noise rms	0.17	°/s

sures yaw rotation. The IDG-300 [94] consists of two vibrating MEMS gyroscopes with capacitive sensing of the movement induced by the Coriolis force. The ADXRS300 [35] is a MEMS vibrating gyroscope with capacitive sensing of the movement induced by the Coriolis force. Tables 5.5 and 5.6 provide the technical characteristics of the IDG-300 and the ADXRS300 as specified by the manufacturers taking into account the specific circuitry of the sensor unit. Sensitivity is expressed in least significant bits (LSB) based on the characteristics of the ADC.

5.4.2. Host unit

The host unit provides power to the sensors, synchronizes the measurements, and handles the communication with the computer. Measurements are triggered by the host so that the sensors measure at the same moment in time. The automatic bus arbitration performed by the CAN module on the AT90CAN128 microcontrollers resolves bus conflicts when the sensors start to simultaneously send the sensor data to the host unit. Two different host units were developed: a wired and a wireless version. The wired version uses the serial interface RS-232 with a data rate of up to 1 Mbit for communication with the computer. The wireless host (see Figure 5.8) is based on the Gumsitx verdex pro, an embedded Linux platform. The wireless host uses Wi-Fi for communication with the computer.

Table 5.6.: ADXRS300 single-axis gyroscope properties

Parameter	Value	Unit
Measurement range	±300	°/s
Sensitivity	1.6	°/s
Nonlinearity	0.1	% of full scale
Bandwidth	400	Hz
Noise rms	2.5	°/s

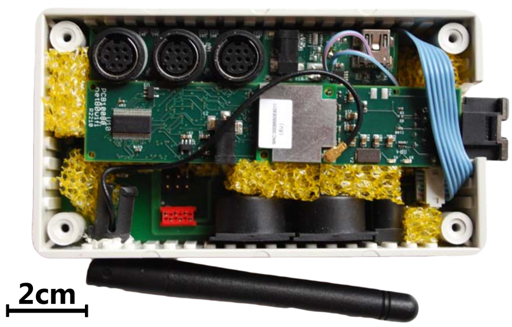


Figure 5.4.: The wireless MotionNet host [149]

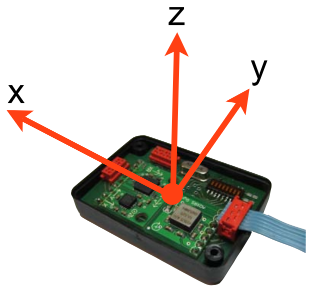


Figure 5.5.: The sensor coordinate system

5.5. Capturing piano performance

Measurement axes: A MotionNet sensor unit measures angular rate and linear acceleration around respectively along the three axes, x, y, and z shown in Figure 5.5. The sign of a angular rate measurement is defined by the right-hand-rule. Let the thumb of the right hand point in the direction of the axis in question. Then, the measured angular velocity has a positive sign if the rotation is in the direction of the fingers. The acceleration measurements along the x-, y-, and z-axis are called x-, y-, and z-acceleration. The angular rates around the x-, y- and z-axis are called x-, y-, and z-rate.

Sensor placement: The sensors are attached to the upper arm, the forearm, and the back of the hand (see Figure 5.6). The upper arm sensor is strapped to the side of the upper arm. The direction of the x-axis is aligned with the direction of the upper arm bone. The forearm sensor is strapped to the forearm near the wrist joint. The sensor on the back of the hand is attached to a thin glove, which is worn by the player. The player's fingertips are free of cloth so that the glove does not interfere with the player's sensor of touch. Furthermore, there is no cloth between the fingers, i. e., the fingers are entirely free, to avoid sensing artifacts that would otherwise occur due to finger movements. To diminish vibrations that can occur at the sensor on the back of the hand, the sensor is additionally stabilized with a cloth from above (see Figure 5.7).

Measurement range: To evaluate whether the measurement range of the selected sensing elements fits well, the movements of the right hand and arm when playing the first of the "Six dances in Bulgarian rhythm" from volume 6 of the Mikrokosmos by B. Bartók were recorded. Table 5.7 shows the maximum and minimum values as well as the 0.1 and 99.9 percentiles. For the sensors worn on the upper arm and the wrist, the maximal and minimal measurements are within the typical measurement range. However, the measurements on the back of the hand can exceed the measurement range in rare cases. In such cases, the minimal values exceed the x- and y-rate measurement range

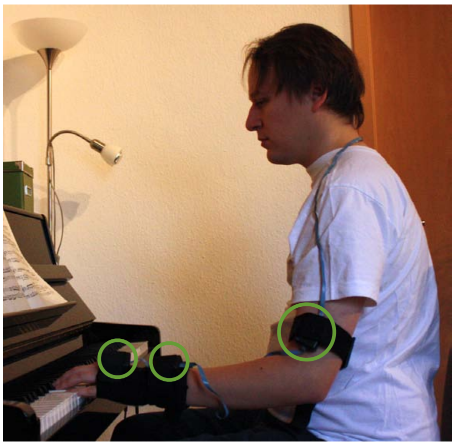


Figure 5.6.: Placement of the sensors on the arm

Figure 5.7.: The sensor on the back of the hand is attached to a glove and stabilized with a cloth from above.

Table 5.7.: Characteristics of playing movements

	Min	0.1 %	99.9 %	Max	Unit
Upper arm x-accel.	− 0.80	− 0.51	+ 1.26	+ 1.36	g
Upper arm y-accel.	+ 0.48	+ 0.60	+ 1.39	+ 1.69	g
Upper arm z-accel.	− 0.59	− 0.46	+ 0.49	+ 0.67	g
Upper arm x-rate	− 69	− 54	+ 89	+ 127	°/s
Upper arm y-rate	− 156	− 106	+ 144	+190	°/s
Upper arm z-rate	− 89	− 72	+ 68	+ 79	°/s
Forearm x-accel.	− 1.29	−1.04	+ 1.80	+ 2.52	g
Forearm y-accel.	− 1.26	− 1.00	+ 1.28	+ 1.93	g
Forearm z-accel.	− 0.99	− 0.66	+ 2.72	+ 3.30	g
Forearm x-rate	− 192	− 137	+ 109	+ 174	°/s
Forearm y-rate	− 240	− 176	+ 174	+ 303	°/s
Forearm z-rate	− 143	− 122	+ 98	+ 116	°/s
Hand x-accel.	− 2.38	− 1.92	+ 1.87	+ 2.56	g
Hand y-accel.	− 2.74	− 1.91	+ 3.54	**+ 4.80**	g
Hand z-accel.	− 2.25	− 1.33	+ 3.28	**+ 3.80**	g
Hand x-rate	**− 619**	− 387	+ 298	+ 388	°/s
Hand y-rate	**− 544**	− 326	+ 311	**+ 548**	°/s
Hand z-rate	− 299	− 212	+ 170	**+ 306**	°/s

and the maximal values exceed the y- and z-acceleration as well as the y- and z-rate[8] measurement ranges. This however only occurs rarely as shown by the 0.1 and the 99.9 percentile, which are always contained in the measurement range. There is an inherent trade-off between increasing sensitivity and measurement range. The sensor has to be sensitive so that the small movements of the upper arm can be sensed and provide adequate measurement range to sense the movements of the more agile hand. As shown, the MotionNet sensor provides a good compromise between these goals although a slightly larger measurement range would be beneficial to sense hand movement more accurately.

5.6. Summary

To be usable for piano pedagogy, a sensor system has to be sensitive, unobtrusive, and inexpensive. Current options to record finger movements, bend-sensing gloves and optical motion capture, are inadequate for piano pedagogy as they are either too obtrusive or prohibitively expensive. The main technologies to sense arm movements are opti-

[8] Note that z-rate measurement is done with the ADXRS300 gyroscope, which has a smaller measurement range of 300°/s than the IDG-300 used for the other axes.

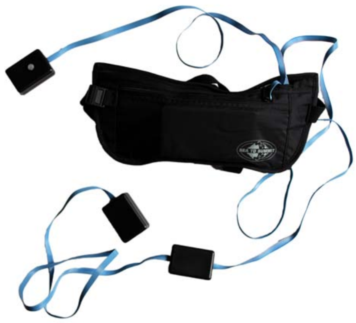


Figure 5.8.: An assembled MotionNet system [149]

cal motion capture, electromagnetic motion capture, electromechanical motion capture, Computer Vision based on consumer cameras, and inertial sensing. Inertial sensing provides the best compromise between sensitivity, unobtrusiveness, and cost. The closest competitor is optical motion capture, which is more unobtrusive but also prohibitively expensive for fast adoption in piano pedagogy. Since commercially available inertial sensors are also not suitable due to limited wearing comfort, too little communication bandwidth for achieving high sampling rates, and cost, it was necessary to develop our own inertial sensing system called MotionNet.

MotionNet (see Figure 5.8) was designed to capture pianist arm movements. It consists of several sensor units, splitters, and a host unit. The sensor units provide 3D linear acceleration and 3D angular rate measurements. The communication between the sensor units and the host unit, which is also worn on the body, is based on a wired CAN-bus. Two different host units were developed, which provide wireless or wired transmission of the data to a computer. It was shown that MotionNet provides a good compromise between measurement range and sensitivity for capturing pianist movements.

Chapter 6.

Pianist movement analysis

Contents

In this chapter, the movement analysis methods introduced in Chapter 4 are used to analyze pianist arm movements. This serves two purposes:

- As already discussed in Section 1.1, it is necessary to distinguish between primary and secondary movement to provide high quality feedback in tasks with a significant amount of secondary movement. Piano playing is such a task: The primary arm movements in piano playing are often small; the secondary movements are relatively large due to key reaction forces. The results obtained in this chapter are later used to realize pedagogical applications (Chapter 8).

- Pianist arm movements are challenging to analyze (as explained in Section 1.2). By showing that our methods perform well here, strong evidence is provided that they are similarly usable to analyze other motor tasks that show a significant amount of secondary movement.

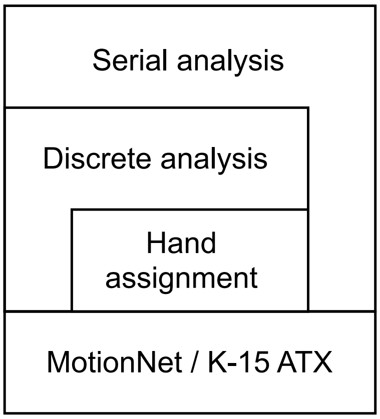


Figure 6.1.: Architecture of the pianist movement analysis system

First, the architecture of the resulting analysis system and the relationship between discrete and serial pianist movement analysis and the other parts of the system presented in this thesis are outlined (Section 6.1). Section 6.2 shows how discrete analysis can be used to detect primary movement when executing a single touch. An evaluation shows the effectiveness of discrete analysis (Section 6.3). Section 6.4 describes the application of serial analysis to detect primary movements that occur over several successive notes. An evaluation demonstrates the effectiveness of serial analysis (Section 6.5).

6.1. Architecture

An overview of the architecture is provided in Figure 6.1. The analysis is based on data provided by the MotionNet sensors (see Chapter 5). Furthermore, MIDI data from a Kawai K–15 ATX upright piano with MIDI interface is used to estimate the key reaction force. However, in order to know on which arm this key reaction force acts and produces secondary movement, it is necessary to determine which hand has pressed a key. For this purpose, methods that assign a hand to a pressed key are necessary. Such hand assignment methods are introduced in Chapter 7. Based on the data from the MotionNet sensors and the estimated key reaction force, discrete analysis estimates the amount of secondary movement and decides whether a primary movement has occurred. The estimation of secondary movement performed by the discrete analysis is passed on to the serial analysis, which allows estimating the overall amount of secondary movement that has occurred during a time interval encompassing several discrete analyses, i. e., several touches. Based on this estimation of overall secondary movement, serial analysis determines whether a primary movement has occurred during a time interval containing several touches.

6.2. Discrete analysis of single touches

In the following discrete analysis is applied to the analysis of movements to execute a single touch. The following topics will be discussed:

- **Measurement:** The player's movements and the factors that influence secondary movement have to be measured (Section 6.2.1).
- **Data collection:** A representative data set of secondary movement has to be collected (Section 6.2.2).
- **Training:** From the possibilities discussed in Section 4.2.1, one has to choose a specific analysis variant and perform training with one part of the collected data set (Section 6.2.3).
- **Evaluation:** Using the other part of the data set, the accuracy of the separation between primary and secondary movement is determined (Section 6.3).

6.2.1. Measurement

Arm movements: To cover all movements of the arm, the seven main degrees of freedom are measured. These are:

1. Hand abduction and adduction
2. Hand extension and flexion
3. Forearm pronation and supination
4. Forearm extension and flexion
5. Upper arm abduction and adduction
6. Upper arm extension and flexion
7. Upper arm rotation

The angular displacement that has occurred during a certain time interval in one of the degrees of freedom can be computed from the corresponding angular velocity signal. The angular velocities are determined based on three MotionNet sensors per arm as shown in Figure 5.6 on p. 71. The angular velocities are denoted $F_i(t)$, where t is time and $i = 1 \dots 7$ indicates one of the seven degrees of freedom (see Table 6.1).

To calculate $F_1(t)$ to $F_4(t)$, the signals of two adjacent sensor units are used. The angular rates in the wrist joint, $F_1(t)$ and $F_2(t)$, are computed as the difference of the corresponding angular rates of the sensor on the hand and on the forearm. The forearm rotation rate $F_3(t)$ is determined as the difference between corresponding angular velocity of the forearm sensor and the upper arm sensor. Upper arm movements $F_5(t)$, $F_6(t)$, and $F_7(t)$ are determined from the upper arm sensor alone, assuming that the

Table 6.1.: Association between i and the seven degrees of freedom of the arm. The ordering of the movement directions (abduction-adduction, extension-flexion, etc.) defines the sign of the corresponding angular rate signal $F_i(t)$. E. g., $F_2(t) > 0$ would that the hand was extended in the wrist; $F_2(t) < 0$ would indicate that the hand was flexed. The rationale for the ordering is to have negative values for movements that bring the fingertip closer to the key and positive for movements that bring the finger away from the key. For abduction-adduction and rotation movements the direction was chosen arbitrarily.

i	Degree of freedom
1	Hand abduction-adduction
2	Hand extension-flexion
3	Forearm supination-pronation
4	Forearm flexion-extension
5	Upper arm abduction-adduction
6	Upper arm flexion-extension
7	Upper arm rotation

upper body remains still. To determine the angular rate of forearm extension and flexion in the elbow joint, it is first necessary to determine the angle of forearm rotation in relation to the upper arm. For this purpose, the posture of the upper arm and the forearm with respect to gravity is determined with two one-dimensional Kalman filters. The corresponding signals on the forearm sensors are then trigonometrically weighted according to the forearm rotation angle. Forearm extension-flexion angular rate $F_4(t)$ is then determined as the difference between the trigonometrically weighted forearm sensor signals and the corresponding angular rate on the upper arm sensor.

The angular displacement in the i-th degree of freedom during a touch movement is calculated with

$$F_i = \int_{t_0-L}^{t_0} F_i(t)\,dt, \tag{6.1}$$

where t_0 is the point in time when the MIDI note-on event was received and L is the length of the discrete analysis interval. The actual value of L was chosen empirically (see Section 6.3).

Estimation of key reaction force: Secondary movement depends on the reaction force, which acts between key and finger. Therefore, key reaction force is an important factor that can be used to estimate secondary movement. We use a method similar to Wolf et al., who use the reported loudness from a MIDI interface to estimate key reaction force [181]. See Section 2.2.2 for a discussion of alternatives.

The largest amount of key reaction force is generated when the key is abruptly stopped by the felt on the keybed when the key is fully depressed. The amount of key reaction

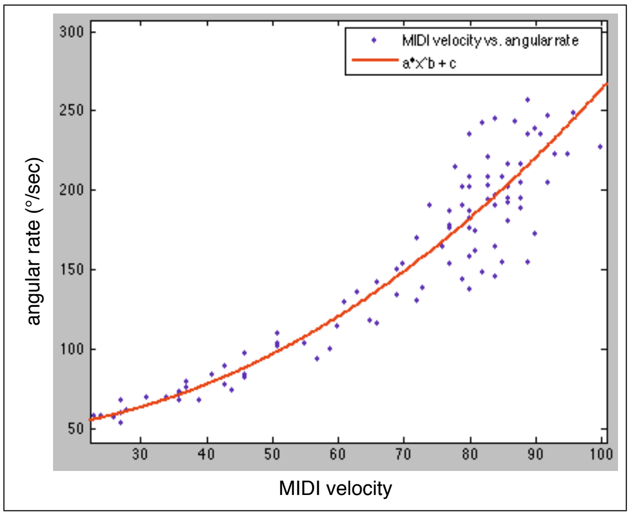


Figure 6.2.: MIDI velocity vs. angular rate.

force that is generated due to this impact depends on the amount of kinetic energy that has to be absorbed. The kinetic energy depends quadratically on the velocity of and linearly on the moving mass ($K = {}^1\!/_2\, m \cdot v^2$). The loudness reported by the MIDI interface allows determining the velocity that is present at the fingertip, which is equal to the velocity of the key. However, the mass cannot be determined from the loudness information alone, i. e., it is not possible to know whether a finger (small mass) or the entire arm (large mass) was used. This is a principle disadvantage of estimating key reaction force based on MIDI data.

The velocity of the key is estimated based on MIDI data from a Kawai K–15 ATX acoustic piano with MIDI interface. For each pressed key, the K–15 ATX reports an integer number that represents the loudness of the generated sound. To determine the mapping from reported loudness to key speed, several touches were recorded. These touches were performed with isolated hand movement from the wrist alone without active participation of the rest of the arm or the fingers. The maximum angular rate of the hand movement was determined for each touch. An exponential function of the form $g(l) = al^b + c$ was fitted to the recorded data using the least mean squares (see Figure 6.2), where l is the measure of loudness reported by the Kawai K–15 ATX. By evaluating the learned function g for a given measure of loudness, the angular rate of the wrist movement to play such a sound can be determined. This value is proportional to the speed of the key. Kinetic energy depends quadratically on velocity and linearly on the moving mass ($K = {}^1\!/_2\, m \cdot v^2$). To determine a value F_K that is proportional to kinetic energy the value of $g(l)$ has to be squared, i. e., $F_K = g(l)^2$. F_K is proportional to kinetic energy under the assumption that the moving mass does not change, which (as already mentioned previously) is a principle weakness when estimating key reaction force based on MIDI data.

6.2.2. Data collection

To learn to estimate the conditional density of secondary movement for the i-th degree of freedom, which is denoted as $p_{si}(y \mid x)$, a representative sampling of this density is needed. Therefore, a data set S_i with touches has to be collected that were executed **without** primary movement in the i-th degree of freedom. A sample $s \in S_i$ composed of the measurement F_i and a vector x of factors that have an influence on secondary movement. Since F_i contains no primary movement the sample have the form

$$s = (F_i, x) = (M_{si} + E_i, x),$$

where M_{si} is the secondary movement and E_i is the measurement error in the i-th degree of freedom.

To collect a representative sampling of $p_{si}(y \mid x)$ it is necessary to systematically vary factors, both measured and non-measured, that may have an influence on secondary movement. An extensive data set of approximately 18.000 touches was recorded. The following factors were systematically varied:

- Loudness: With increasing loudness, the key-reaction forces increase leading to greater amounts of secondary movement. Touches were executed in three loudness bands: pianissimo to mezzo piano (pp–mp), mezzo-piano to mezzo-forte (mp–mf), and mezzo-forte to fortissimo (mf–ff).

- Mass: A greater mass traveling with the same velocity has a greater kinetic energy, which results in higher forces when the mass is stopped at an impact. Therefore, the mass involved in the execution of the movement was altered. Touches were executed with isolated movement of the fingers, the hand, the forearm (extension in the elbow), forearm rotation, or arm extension in the shoulder joint. For convenience, the touches will be called finger touches, hand touches, forearm touches, pronation touches, supination touches, and shoulder touches.

- Finger: The position from where the key-reaction force acts on the arm influences secondary movement. Therefore, the finger that executes the touch was varied.

- Struck and pressed touches: Struck and pressed touches are known to have different key-reaction force profiles [81, 106] (see Section 2.2.2). In consequence, it is sensible to assume that the type of touch has an influence on secondary movement. Therefore, struck and pressed touches were executed.

- Flexion- and extension touches: Each finger has three joints. To press down a key, flexion movement can be used in all three joints [123], which we will call a flexion-touch, or the finger is only flexed in the knuckle joint while it is extended in the other two joints [123], which we will call an extension-touch. Both flexion- and extension touches were executed.

Altogether, the different loudness levels (3), touch types (6), fingers (5), and direct vs. indirect touch (2) results in 180 combinations (3·6·5·2). For each combination, a session of approximately 100 samples was recorded resulting in an overall of approximately 18.000 samples. For finger touches, the 100 samples were split into 50 extension- and 50 flexion touches. The entire data set is denoted S. The data sets $S_i \subset S$ contain only the samples that were executed without primary movement in the i-th degree of freedom.

6.2.3. Training

Now that data sets S_i with samples of secondary movements are available, it is necessary to select a specific analysis variant from the options to estimate $p_{si}(y \mid x)$ discussed in Section 4.2.1. The decision process sketched in Figure 4.2 on p. 57 helps to make an informed decision. For analyzing piano playing movements, high sensitivity is needed so that recognition of primary movements based on thresholding is inadequate. Since we want to use serial analysis to analyze movements that span several successive touches (see Section 6.4), quantile regression is not adequate. The remaining options are maximum likelihood estimation and heteroscedastic regression.

Using a parametric model has two advantages. First, it allows us to model the dependencies of secondary movement from the influence factors based on kinesthetic insight and guided experimentation, which can help to achieve good estimation results. Second, it reduces the amount of training data that has to be provided since the model is already provided and does not have to be learned from the data set. For these two reasons, we chose to use a parametric model. Both maximum likelihood estimation and heteroscedastic regression can be used with a parametric model. We chose to use maximum likelihood estimation as this provides the flexibility to choose the density type of $p_{si}(y \mid x)$ freely.

To perform maximum likelihood estimation, it is first necessary to choose the density type of $p_{si}(y \mid y)$ and a parameterized function f_i that is used to determine the parameters of the density. We chose to model the density of $M_{si} + E_i \mid x$ as a normal distribution:

$$M_{Si} + E_i \mid x \sim \mathcal{N}(\mu_i, \sigma_i),$$

where the mean μ_i and standard deviation σ_i is computed by the function f_i from the factors of influence x:

$$(\mu_i, \sigma_i) = f_i(x).$$

The function f_i is parameterized. Its concrete form was learned using maximum likelihood estimation. Two variants of the function f_i were used:

- Variant 1, called the minimal model, computes the mean μ_i and standard deviation σ_i as function of F_K, the estimation of key reaction force, i. e., $(\mu_i, \sigma_i) = f(F_K)$.

- Variant 2, called the full model, computes the mean μ_i and standard deviation σ_i as function of F_K and the movement measurements in the other degrees of freedom, i. e., $(\mu_i, \sigma_i) = f(F_K, F_1, \ldots, F_{i-1}, F_{i+1}, \ldots, F_7)$.

Minimal model: The formulas for the minimal model are

$$\mu_i = \alpha_i F_K$$
$$\sigma_i = \beta_i F_K + \gamma_i.$$

The mean μ_i is modeled as a linear function of F_K without an intercept. As the amount of key reaction force increases, the arm tends to be displaced more and more in one direction. Consequently, the mean μ_i moves away from $\mu_i = 0$ as F_K increases. In the limit $F_K = 0$ the mean μ_i should be zero since, although small unintended movements cannot be completely avoided, no forces are present that could systematically displace the arm in one direction. Therefore, the computation of μ_i does not include an intercept. The standard deviation σ_i is modeled as a linear function of F_K with an intercept. When F_K increases, the spread of secondary movement increases, too, which leads to an increasing σ_i. Since small unintended movements cannot be completely avoided, there is always some spread. Therefore, the computation of σ_i contains an intercept.

Full model: The formula for the full model is

$$\mu_i = \alpha_i F_2^- + \beta_i F_3^p + \gamma_i F_3^s + \delta_i F_4^- + \epsilon_i F_6^- + \zeta_i F_K$$
$$\sigma_i = \eta_i F_2^- + \theta_i |F_3| + \iota_i F_4^- + \kappa_i F_6^- + \lambda_i F_K + \nu_i.$$

The term F_i^- denotes the downward movement towards the key performed by the i-th degree of freedom. It is computed as

$$F_i^- = \begin{cases} -F_i & \text{if } F_i < 0, \\ 0 & \text{else.} \end{cases}$$

The term F_3^p denotes the forearm pronation movement and the term F_3^s denotes the forearm supination movement. They are calculated as

$$F_3^p = \begin{cases} -F_3 & \text{if } F_3 < 0, \\ 0 & \text{else,} \end{cases} \qquad F_3^s = \begin{cases} F_3 & \text{if } F_3 > 0, \\ 0 & \text{else.} \end{cases}$$

By definition F_i^-, F_3^p, and F_3^s are always greater than or equal to zero. The terms $\zeta_i F_K$ used in the computation of the mean μ_i and $\lambda_i F_K + \xi$ used in the computation of the standard deviation σ_i correspond to the minimal model. The secondary movement in the i-th degree of freedom may not depend on the measurement of the movement in the i-th degree of freedom F_i. Therefore, when a function f_i where is is one of $i = 2, 3, 4, 6$ the corresponding variable or variables F_2^-, F_3^p, F_3^s, F_4^-, or F_6 are set to zero.

Maximum likelihood estimation: In the following maximum likelihood estimation [12, p. 23] is applied to determine the values of the parameters α_i to ν_i of the function f_i.

The likelihood of the data set S_i given f_i is

$$p(D \mid f_i) = \prod_{j=1}^{l} p(F_i(j) \mid x(j)) = \prod_{j=1}^{l} \mathcal{N}\big(F_i(j) \mid (\mu_i, \sigma_i) = f_i(x(j))\big),$$

where $F_i(j)$ denotes the movement measurement for the j-th of l samples in the data set S_i and $x(j)$ denotes the factors of influence in the j-th sample in the data set. To avoid underflow problems the logarithm is maximized

$$\ln p(D \mid f_i) = \sum_{j=1}^{l} \mathcal{N}\big(F_i(j) \mid (\mu_i, \sigma_i) = f_i(x(j))\big).$$

The maximum likelihood estimate of the parameters $w = (\alpha_i, \beta_i, \gamma_i)$ for the minimal model or $w = (\alpha_i, ..., \nu_i)$ for the full model is given by

$$\underset{w}{\text{maximize}} \sum_{j=1}^{l} \mathcal{N}\big(F_i(j) \mid (\mu_i, \sigma_i) = f_i(x(j); w)\big).$$

A constraint nonlinear optimization algorithm was used to find the parameters by maximizing this expression. The calculated value of the standard deviation σ_i has to be always greater than zero. To assure this, the maximization is done under certain constraints: For the minimal model it is required that $\beta_i \geq 0$ and $\gamma_i > 0$; for the full model it is required that $\eta_i \geq 0$, $\theta_i \geq 0$, $\iota_i \geq 0$, $\kappa_i \geq 0$, $\lambda_i \geq 0$, and $\nu > 0$. Since the F_i^- and F_K are also always greater than or equal to zero, the computed standard deviations σ_i are always positive.

6.3. Evaluation of discrete analysis

The experimental results presented in this section address questions that have remained open in the preceding discussion. First, the analysis with the minimal model is compared to analysis with the full model (Section 6.3.1). It is shown that the estimation quality is better when using the full model. This shows that the movements in the other joints of the arm provide valuable information for estimating secondary movement. Second, the accuracy of primary movement detection is presented in Section 6.3.2. Finally, it is examined how much training data is necessary to train the full model (Section 6.3.3).

6.3.1. Estimation quality

Minimal model: Figure 6.3 compares the estimation results obtained by the minimal model with the measurements of actual secondary movements in the data set. F_6, which is the measurement of arm movement from the shoulder, is indicated on the y-axis. F_K, which is calculated from the loudness reported by the MIDI interface of the Kawai K–15 ATX, is indicated on the x-axis. The solid lines show the estimated mean μ_6 (middle line) and standard deviation $\mu_6 \pm \sigma_6$ (outer lines) computed by the minimal model. The plotted data set contains all touches excluding the touches that were performed with primary movement from the shoulder (shoulder touches). However, it can be seen in Figure 6.3, the density of secondary movement and error depends not only on the key velocity F_K but also on what movement was used to execute the touch. The different touch types in the data set are represented with different colors.

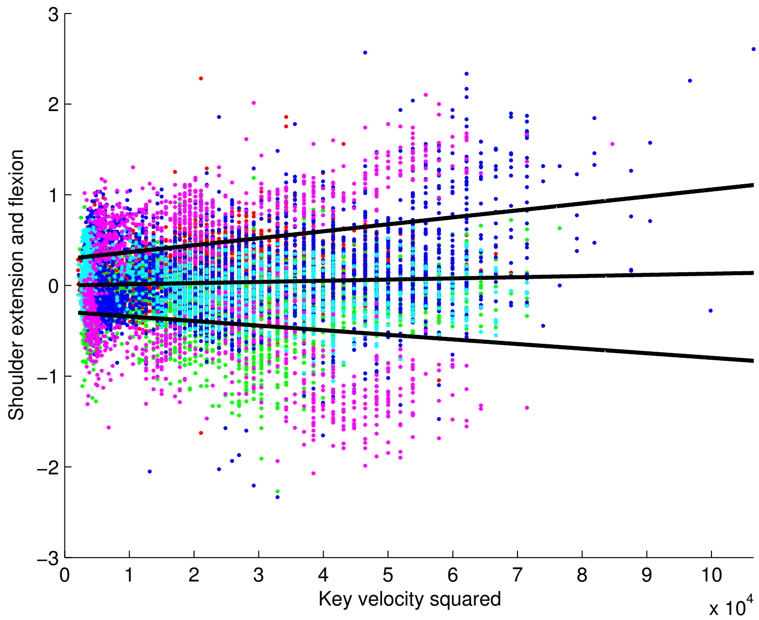


Figure 6.3.: Mean and standard deviation estimation (minimal model).

Full model: The full model on the other hand uses movement measurement in addition to F_K to estimate the density of secondary movement and error. This improves the estimation of secondary movement. Evidence for this is provided by Figures 6.4 and 6.5. Similar to the Figure 6.3, the x-axis indicates the measured key velocity F_K and the y-axis, the measured movement of the arm in the shoulder joint F_6. In contrast to Figure 6.3, Figures 6.4 and 6.5 contain only one type of touch movement: Figure 6.4 contains only finger touches, i. e., elements of the data set D_0 and Figure 6.5 only elbow touches, i. e., elements of the data set D_4. As F_K is computed from a discrete MIDI velocity signal, it is possible to split the data into sets of samples with exactly the same value F_K. Each of these sets corresponds to one vertical slice of the graph. For each vertical slice, the average mean and the average standard deviation is computed by evaluating f_i for each sample. A sample is indicated with a tiny dot while the average mean and the average mean ± the average standard variation of a slice are marked with thicker dots of black respectively blue color. By examining the data points and the estimated means and standard deviations in Figures 6.4 and 6.5, it is evident that the estimations are sensible. Furthermore, the full model's estimation adapts to the performed type of touch (see Figures 6.5 and 6.4). Note that the full model had no information about the performed touch but was able to find the estimates by evaluating f_i. The complete set of figures similar to Figures 6.3 and 6.4 were obtained for all seven degrees of freedom and for all touch types. They can be found in Appendix A.

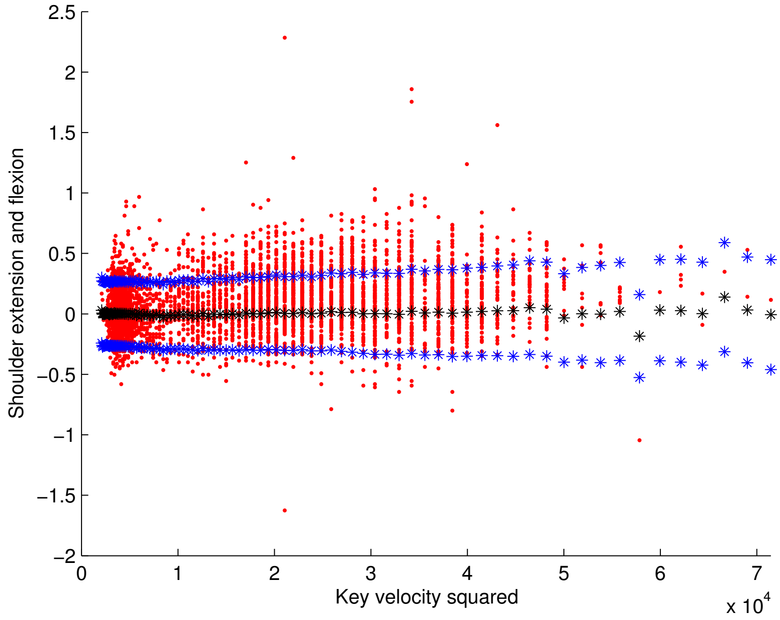


Figure 6.4.: Estimation for finger touches (full model)

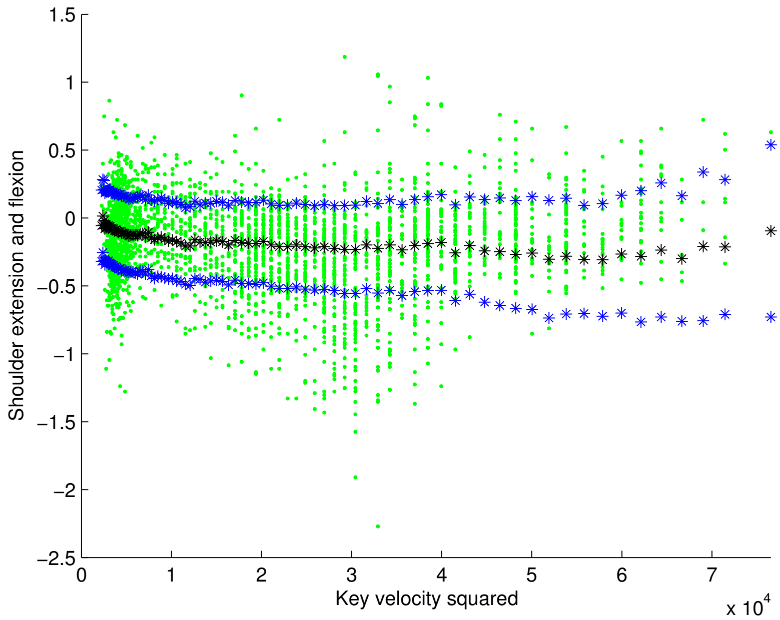


Figure 6.5.: Estimation for forearm touches (full model)

Primary movement detection vs. d: As discussed in Section 4.2.2, a primary movement is detected if the measurement, in this case F_i, is unlikely with respect to the conditional density of secondary movement, i. e., if $p_{si}(y = F_i \mid x) < c$, where c is a constant. As $p_{si}(y \mid x)$ is modeled as a normal distribution, a constant d can be found so that

$$p_{si}(y = F_i \mid x) < c \iff |F_i - \mu_i| > d \cdot \sigma_i. \tag{6.2}$$

By examining graphs of primary movement detection vs. the constant d (such as the graphs depicted in Figure 6.6), it is possible to assess the estimation quality. The y-axis indicates in how many percent of the cases, a primary movement is detected. The x-axis indicates the value of d. The different colors in the diagrams represent touches with different primary movements: These correspond to the different touch types in the data set. The graphs in Figure 6.6 show the detection of primary wrist extension and flexion movement. The estimation quality can be judged by two properties:

- A good estimation shows a slow decay of the graph of the touch with primary movement in the examined degree of freedom. This ensures that a large value can be used for d so that only a small number of false positive detections occur. For the top row of Figure 6.6, this means that the green graph, which represents hand touches, should decay slowly.
- All other graphs should decay quickly since they do not contain primary movement in the examined degree of freedom. This ensures that a small value for d can be used so that only a small number of false negative detections occur.

By examining the diagrams in Figure 6.6 in this way, one can see that the estimation quality of the full model is better than the estimation quality of the minimal model.

6.3.2. Detection of primary movement

In order to detect primary movement, it is first necessary to choose the constants d and L. The constant d (see Equation 6.2) weighs between false positive and false negative detections. The constant L (see Equation 6.1) determines the length of the analysis interval.

Choosing d: The constant d controls the sensitivity of the primary movement detection. If the chosen value of d is too high, small primary movements are not detected (false negative). However if the chosen value of d too low, secondary movements can be mistaken for primary movement (false positive). The percentage of false positive detection decreases as the value of d increases. The value of d is determined using a cost function that punishes false positive and false negative detections equally

$$\mathrm{cost_i} = \frac{1}{2}\left(\frac{\text{count(false positive)}}{|S_i|} + \frac{\text{count(false negative)}}{|S \setminus S_i|}\right). \tag{6.3}$$

The factor $1/2$ is used to normalize the cost function to the interval $[0, 1]$.

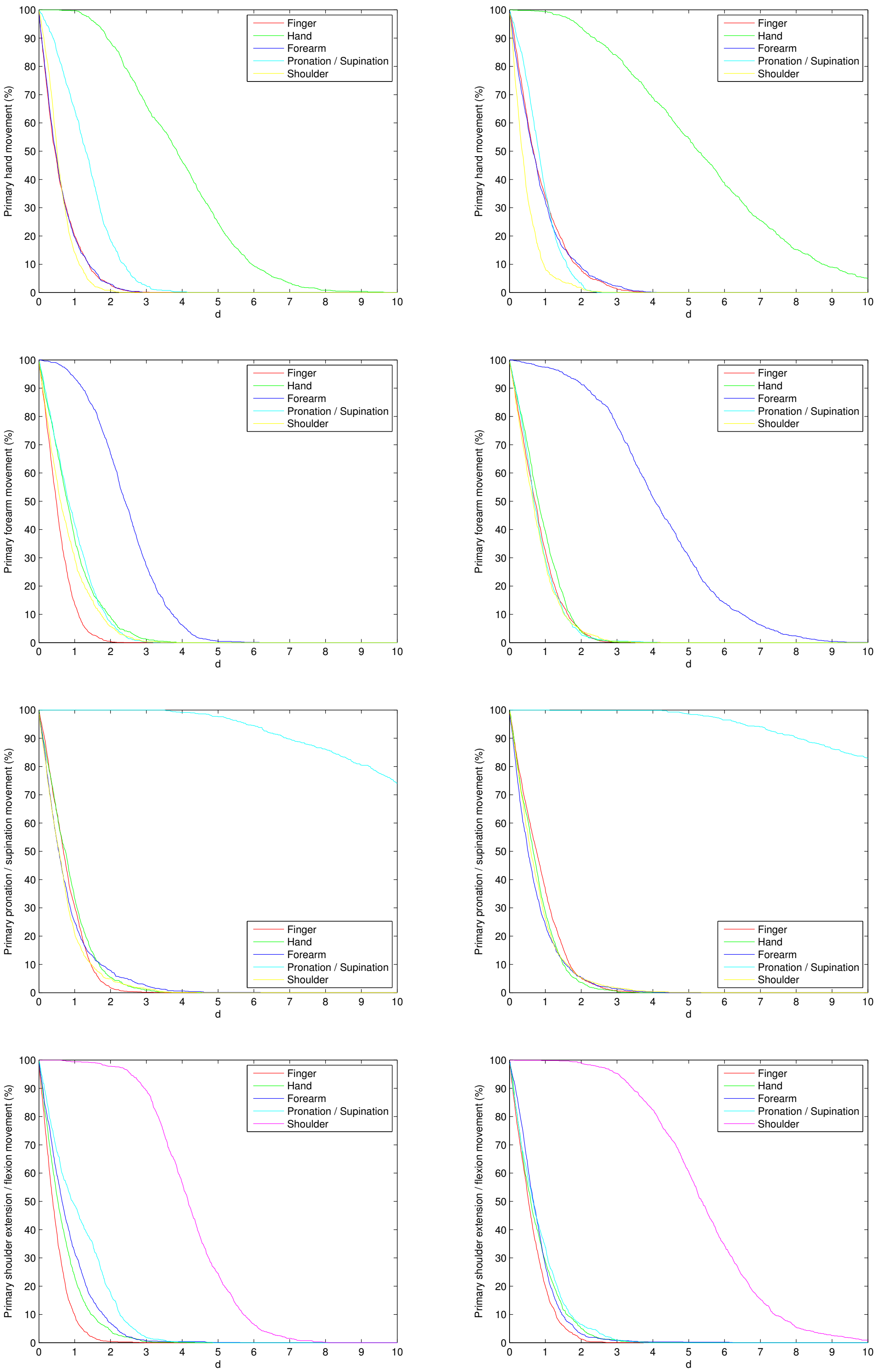


Figure 6.6.: Detection of primary movement vs. d: The left side shows the analysis with the minimal model. The right side shows the analysis with the full model. (The graphs were drawn based on a randomized selection of one third of the data, which was reserved for testing and not used for training.)

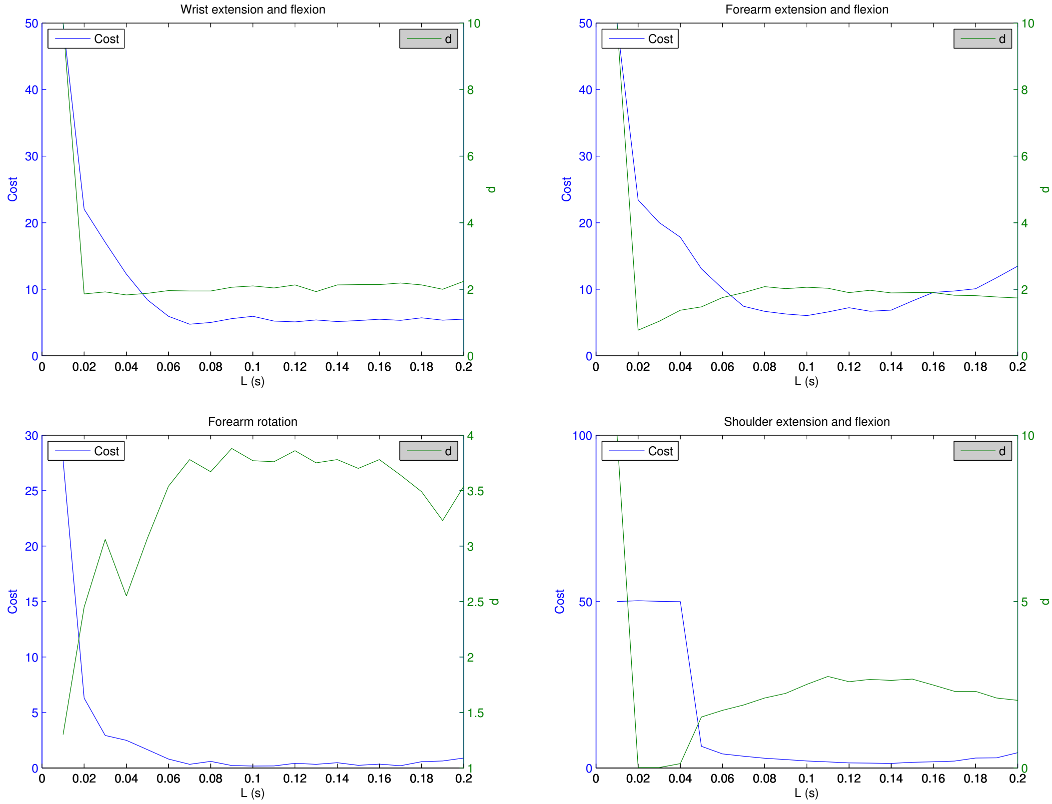


Figure 6.7.: Cost vs. L: The green graph indicates the optimal value for d (primary axis). The blue graph indicates the resulting cost (secondary axis). Cost is expressed in %. (The cost was determined based on a randomized selection of one third of the data, which was reserved for testing and not used for training.)

Choosing L: Using this cost function, it is possible to determine the optimal value for L, the length of the discrete analysis interval (see Equation 6.1). For a given L the optimal value of d is determined. The associated cost is stored. Doing so reveals that starting with the smallest possible value of L (which is 0.01 s as the sensors were sampled at 100 Hz) the accuracy increases with increasing L (see Figure 6.7). Since at values larger than $L = 0.1$ s only small accuracy changes are visible and a small value of L is beneficial when analyzing quick movements, the value $L = 0.1$ s was chosen. Consider, e. g., a series of quick hand touches to execute chord repetitions. If the value of L was chosen too large, the upward movement of the hand to release the keys would cancel out the downward movement in the calculation of F_2 (see Equation 6.1). This would make primary movements harder to detect and could lead to false negative detections.

Detection accuracy: Six types of touch movements are contained in the data set S: finger touches, hand touches, forearm, touches, pronation touches, supination touches and shoulder touches. Hand touches contain primary wrist flexion movement. Forearm touches contain primary elbow extension movement, pronation and supination touches contain primary forearm rotation movements. Shoulder touches contain primary shoulder extension movement. Table 6.2 shows the accuracy of primary movements detection. To determine the accuracy of primary movement detection, the optimal value of d with respect to the cost function (see Equation 6.3) was determined first. The true positive detection rate for primary movements varies from joint to joint between 91.16 and 99.87%. False positive detections range from 0 to 7.83%. The best detection quality is achieved for forearm rotation movements. Primary forearm rotation movements are detected with a rate of 99.87 (true positive). False positive detections of primary forearm movement are very rare (0 to 0.52% depending on the used touch type). The second best rates are achieved for the detection of primary shoulder extension/flexion movements. Here primary movements are detected in 97.48% of the cases while false positive are made at a rate of 0.35 to 3.28%. Primary wrist and elbow flexion/extension movement is detected with the least accuracy. Primary movement is detected with a rate of 92.36 (wrist) respectively 91.16% (elbow). Also false positive detections are more common ranging between 0.94 to 7.83% (wrist) and between 2.52 and 3.69% (elbow). The reason for the lower detection accuracy for elbow and wrist movements, is that these movements are particularly small in comparison with the secondary movement that occurs in these joints. While extension/flexion movements in the shoulder joint are also very small, the amount of secondary movement that is typically experienced in the shoulder is lower, which enables better detection rates. The forearm rotation movements are larger in terms of angular displacement, making primary forearm rotation movements easier to recognize.

Implications for piano pedagogy: When providing feedback for piano pedagogy, unjustified corrective feedback is problematic. Two types of errors can be distinguished:

1. **False negative:** The user has to execute a primary movement. He does so but the system does not detect a primary movement and provides (unjustified) corrective feedback. Next time, the user exaggerates the movement.

2. **False positive:** The user is instructed to avoid to perform a primary movement in an arm joint. He does so but the system detects a primary movement and provides (unjustified) corrective feedback. Next time, the user tries to minimize the movement, e. g., by tensing up the muscles.

Exaggerating the movements or tensing up to satisfy the feedback system are problematical since this can disrupt the naturalness of the user's movements. Therefore, it is important to reduce unjustified corrective feedback as much as possible. To minimize the amount of unjustified corrective feedback, it is possible to manipulate the parameter d. To confirm primary movement when this is demanded, a lower value for d can be used; to confirm that no primary movement has occurred, a higher value for d can be used.

Table 6.2.: Confusion matrix for primary movement detection for optimal values of d (indications are in percent). (The confusion matrix was determined based on a randomized selection of one third of the data, which was reserved for testing and not used for training. The parameter d was determined based on the part reserved for training.)

Detected Actual	Wrist (ext./flex.)	Elbow (ext./flex.)	Forearm rotation (pron./supin.)	Shoulder (ext./flex.)
Finger touch	6.74	3.69	0.00	0.35
Hand touch	**92.36**	3.19	0.00	1.54
Forearm touch	7.83	**91.16**	0.30	1.51
Rotation touch	1.51	2.52	**99.87**	3.28
Shoulder touch	0.94	3.66	0.52	**97.48**

This is not necessary for analyzing forearm rotation movement as the accuracy is very good at the optimal value of d but advisable for analyzing wrist, elbow, and (although to a lesser extent) for shoulder extension and flexion movement.

6.3.3. Size of the training set

To assess the influence of the amount of training data on the goodness of the learned estimations the amount of training data was varied from 1 to 99% of the available data. The rest of the data was used for testing. The optimal value for d was determined as discussed previously. Figure 6.8 shows the value of the cost function and the optimal value of d in dependence of the size of the training set. It is evident that the cost function always remains in a narrow band. Even when only 1% of the available data is used, i. e., 180 touches, no distinct negative effect on the recognition accuracy is notable. This is important for a user that is interested to train the system himself to achieve the best possible recognition results.

6.4. Serial analysis of successive touches

Serial analysis is used to analyze piano playing movements that span over several successive touches. It is not feasible to tackle this problem with discrete analysis, since this would require collecting a representative data set. However, the dimension of such a data collection effort is prohibitively large since the variation possibilities increase exponentially with the number of notes. Some parameters that would have to be varied are the number of notes in the analysis interval, the loudness of each individual note, the rhythm, the tempo, what primary movements are used, etc.

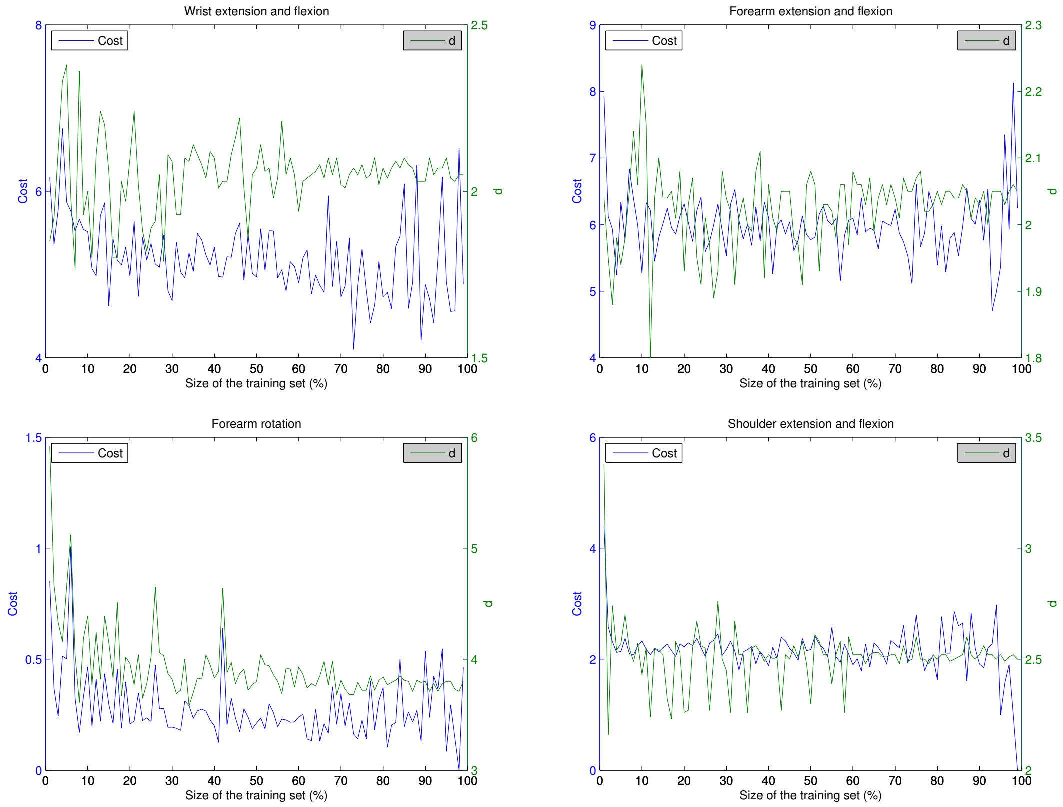


Figure 6.8.: Cost (primary axis) and the optimal value of d (secondary axis) vs. amount of training data. Cost is indicated in %. The cost was determined based on the fraction of the data set not used for training. This also explains the higher variance of the determined cost close to 100% usage of training data as only little testing data is available in this case.)

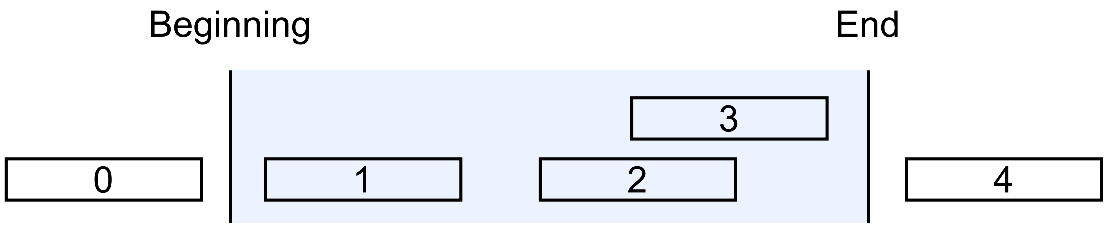


Figure 6.9.: The shaded area indicates the serial analysis interval. The contained discrete analyses are labeled 1, 2, and 3. There is a gap between the discrete analyses 1 and 2. Furthermore, the discrete analyses 2 and 3 overlap.

6.4.1. Combination

Each discrete analysis extends over a certain time interval. As discussed previously, the discrete analysis interval ends when the K–15 ATX reports a note via the MIDI interface and begins $L = 0.1\,\mathrm{s}$ before the MIDI event is received (see Equation 6.1). Serial analysis extends over a larger analysis interval containing several successive touches. The discrete analyses contained in the serial analysis interval can overlap and there can also be a gaps between successive discrete analyses (see Figure 6.9).

Let there be N discrete analyses contained in the serial analysis interval and let $M_{si}(j)$ indicate the secondary movement generated by the j-th touch movement and $E_i(j)$ indicate the measurement error that occurs during the j-th touch movement, then the sum of total secondary movement and measurement error $M_{si,total} + E_{i,total}$ experienced in the analysis interval is

$$M_{si,total} + E_{i,total} = M_{si}(g) + E_i(g) - E_i(o) + \sum_{j=1}^{N} M_{si}(j) + E_i(j),$$

where $M_{si}(g)$, $E_i(g)$, and $E_i(o)$ are terms to compensate for secondary movement and error that occurred in the gaps (g) and and error that was counted several times due to overlaps (o). Since the densities of $M_{si}(g)$, $E_i(g)$, and $E_i(o)$ are unknown as they are not provided by the discrete analysis discussed above, the sum of overall secondary movement and measurement error is approximated with

$$M_{si,total} + E_{i,total} \approx \sum_{j=1}^{N} M_{si}(j) + E_i(j)$$

The probability density of the terms $M_{si}(j) + E_i(j)$ given the factors of influence $x(j)$ of the j-th discrete analysis in the analysis interval is $p_{si}(y \mid x(j))$. As discussed in Section 4.3 the probability density $p_{si,total}(y)$ of $M_{si,total} + E_{i,total}$ can be derived from the conditional densities of secondary movement $p_{si}(y \mid x(j))$

$$p_{i,total}(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ity} \prod_{j=1}^{N} F_t(p_{si}(z \mid x(j)))\ dt.$$

However, since the $M_{si}(j) + E_i(j) \mid x(j)$ are normally distributed there exists a computationally less expensive way to determine $p_{i,total}(y)$.

Restricted boundaries: Let the beginning t^- and the end t^+ of the serial analysis interval be restricted to gaps (e. g., let t^- lie in the gap between discrete analysis 0 and 1 in Figure 6.9 and t^+ between 3 and 4). Furthermore, let $\mu_i(j)$ and $\sigma_i(j)$ be the mean and standard deviation of secondary movement and error for the j-th of N touch movements in the analysis interval $[t^-, t^+]$, i. e., $(\mu_i(j), \sigma_i(j)) = f_i(x(j))$. Then the overall secondary movement and error is distributed with mean

$$\mu_{i,total} = \sum_{j=1}^{N} \mu_i(j)$$

and variance

$$\sigma^2_{i,total} = \sum_{j=1}^{N} \sigma_i(j)^2.$$

Unrestricted boundaries: The limitation that t^- and t^+ have to lie inside gaps can be overcome by spreading the estimation of conditional secondary movement $p_{si}(y \mid x)$ evenly over the time of the discrete analysis interval. Let $b(j)$ denote the beginning and $e(j)$ the end of the j-th discrete analysis interval, which may be completely or partly contained in the serial analysis interval. The density of secondary movement and error $p_s(y \mid x(j))$ for the j-th discrete analysis is evenly distributed over the discrete analysis interval by

$$p_s(y \mid x(j)) = \int_{b(j)}^{e(j)} p_s(y \mid x(j); t)\ dt,$$

where $p_s(y \mid x(j); t_1) = p_s(y \mid x(j); t_2)$ for $t_1, t_2 \in [b(j), e(j)]$. Let $\mu_i(j, t)$ and $\sigma_i(j, t)$ be the mean and the standard deviation of the density $p_s(y \mid x(j); t)$. To satisfy the above equation, the mean $\mu_i(j, t)$ has to have the property that

$$\int_{b(j)}^{e(j)} \mu_i(j, t)\ dt = \mu_i(j)$$

and the variance $\sigma_i(j, t)$ has to have the property that

$$\int_{b(j)}^{e(j)} \sigma_i(j, t)^2\ dt = \sigma_i(j)^2.$$

The mentioned properties are satisfied by defining the mean $\mu_i(j, t)$ as

$$\mu_i(j, t) = \frac{\mu_i(j)}{e(j) - b(j)}$$

and the standard deviation $\sigma_i(j, t)$ as

$$\sigma_i(j, t) = \frac{\sigma_i(j)}{\sqrt{e(j) - b(j)}}$$

if t is inside the interval $[b(j), e(j)]$. Otherwise the mean $\mu_i(j,t)$ and standard deviation $\sigma_i(j,t)$ are set to zero, i. e.,

$$\mu_i(j,t) = \sigma_i(j,t) = 0 \text{ if } t \notin [b(j), e(j)]$$

Now, it is possible to compute the mean

$$\mu_{i,total} = \int_{t^-}^{t^+} \sum_{j=1}^{N} \mu_i(j,t)\ dt$$

and variance

$$\sigma_{i,total}^2 = \int_{t^-}^{t^+} \sum_{j=1}^{N} \sigma_i(j,t)^2\ dt$$

of the overall secondary movement and error $M_{si,total} + E_{i,total}$ in an arbitrarily defined analysis interval $[t^-, t^+]$.

6.4.2. Decision

To decide whether a primary movement was used in the i-th degree of freedom, the mean $\mu_{i,total}$ and the standard deviation $\sigma_{i,total}$ is compared with the measurement of overall movement in the analysis interval, which is computed with

$$F_{i,total} = \int_{t^-}^{t^+} F_i(t)\, dt.$$

A primary movement is detected if the movement measurement $F_{i,total}$ exceeds the mean of the overall secondary movement and error $\mu_{i,total}$ more than a constant d times the total standard deviation $\sigma_{i,total}$

$$|F_{i,total} - \mu_{i,total}| > d \cdot \sigma_{i,total}.$$

The constant d allows to weighing between false positive and false negative errors.

6.5. Evaluation of serial analysis

To determine the accuracy of serial analysis, the proposed method was evaluated based on recorded movement and MIDI data. Arpeggios were performed with or without forearm rotation. Parameters that influence the analysis, the movement, or both were systematically varied:

- **The number of notes:** The number of notes were varied. Four different motifs were played (see Figure 6.10): the first motif contained four notes, the second contained six, and the third motif contained eight notes. Since secondary movement is generated through mechanical interaction with the piano action, a greater number of interactions may lead to a greater amount of secondary movement, which would make primary movement detection more difficult.

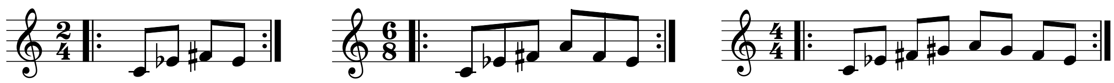


Figure 6.10.: Four-note motif (left), six-note motif (center), and eight-note motif (right)

- **Loudness:** When playing louder, the amount of key-reaction force is increased, which leads to more secondary movement. The motifs were played piano, mezzo-forte, and forte.

- **Tempo:** The tempo has an effect on the primary forearm rotation. When higher tempos are played the rotation is performed with greater speed. Furthermore, the overall size of the movement can be reduced, which makes primary movement detection more difficult. The motifs were recorded at different tempos. The quarter note was played with 60, 100, 140, and 180 beats per minute, which was dictated by a metronome.[1] To generate a recording that produces significant overlaps in the analysis, the motifs were arpeggiated: The highest and lowest note were sustained. The ascending and descending intervals were then played in rapid succession.

The variations in the number of notes, loudness, and tempo, result in 45 combinations. Each combinations was repeated 10 times with and 10 times without forearm rotation so that 900 samples were collected in total.

In the four-note motif, the following movement may occur: The motif begins with the note C. The player begins to supinate shortly after playing C, the lowest note. The notes E-flat and F-sharp are played with supination. Shortly after the note F-sharp, the highest note, is reached, the player reverses the movement direction and plays the notes E-flat and C with pronation. The six- and eight-note motifs are executed similarly. Supination movement may be used when playing ascending intervals, pronation movements when playing descending intervals. The movement direction is reversed when holding the highest or lowest note.

To apply the serial analysis, it is necessary to define the analysis interval $[t^-, t^+]$. The beginning of the analysis interval t^- is halfway between the note C and E-flat. The end of the analysis interval t^+ is the onset time of the highest note. This is the note F-sharp in the four-note motif and the note A in the six- and eight-note motif. For the analysis of the pronation movement, t^- and t^+ were placed correspondingly: The analysis interval starts halfway between the highest and the next note and ends with the onset time of the lowest note. A primary movement is detected if the total movement $F_{i,total}$ exceeds the total mean $\mu_{i,total}$ more than a four times the total standard deviation $\sigma_{i,total}$, i.e., if $|F_{i,total} - \mu_{i,total}| > d \cdot \sigma_{i,total}$ with $d = 4$. This is known to provide a good separation between rotation and non-rotation movements for single touches (see Section 6.3.2). Detection rates of over 96% were achieved (see Table 6.3).

[1] To make it easier for the performer, the six-note motif was played with the metronome indicating dotted quarter notes at 40, 67, 93, and 120 beats per minute, which results in the same temporal distance between two notes like playing with quarter notes with 60, 100, 140, and 180 beats per minute.

Table 6.3.: Accuracy of the serial analysis

Recognized Played	Rotation	No rotation
Rotation	**96.7%**	3.3%
No rotation	3.0%	**97.0%**

Discussion: Serial analysis is not as accurate as discrete analysis. While primary forearm rotation movement detection is possible at a rate of 99.87% when performing discrete analysis with false positive detections of up to 0.52%, the detection rate of primary forearm rotation movement when performing serial analysis is possible only at rate of 96.7% with a rate of false positive detections of 3.3%. There are two reasons for this: First, serial analysis addresses a more difficult problem: When playing several notes, the individual displacements due to key reaction force accumulate so that more secondary movement is generated. The primary movements however do not increase in the same dimension. Second, serial analysis is based on simplifications that introduce estimation error, namely the assumption of statistical independence of the $p_{si}(y \mid x(j))$ discussed in Section 4.3.1. Furthermore, it is assumed that the secondary movement and error contained in the gaps or counted multiple times if they overlap contributes only little to the overall amount of secondary movement.

As discussed in Section 6.3.2, unjustified corrective feedback is highly problematic in context of piano pedagogy. To avoid unjustified corrective feedback, we advise to manipulate the value of d so that a lower value of d is used when the student is instructed to execute a primary movement and a higher value of d is used when the student is instructed to avoid primary movement.

In other application areas, the separation between primary and secondary movement is usually clearer so that similar or better recognition rates can be expected.

6.6. Summary and interpretation of results

Summary: In this chapter, discrete and serial analysis were used to analyze piano playing movements. Discrete analysis was used to analyze a single touch. Serial analysis was used to analyze a series of successive touches. Discrete analysis of a single touch is based on movement measurements in the joints of the arm and on the estimation of key reaction force from MIDI data. To determine the conditional density of secondary movement, an extensive data set containing various movement variants was used to train two alternative functions that model the influences on secondary movement. The first function, the minimal model, estimates the density of secondary movement as a function of MIDI velocity. The second function, the full model, uses the movement measurement in other parts of the arm in addition to MIDI velocity to improve the estimation of secondary movement. An evaluation shows the effectiveness of using discrete analysis for analyzing touch movements. Furthermore, it is shown that the analysis with the full model is better than analysis with the minimal model. This shows that it is beneficial to take into account the movements that are present in the other parts of the arm to estimate secondary movement in a particular joint.

To analyze a series of successive touches, discrete analysis is not applicable since this would require to collect a prohibitively large data set as the possible variations increase exponentially with the number of notes. Serial analysis on the other hand allows analyzing a series of successive touches without additional data collection effort based on the estimation made by the discrete analysis. An evaluation shows that serial analysis is effective to detect primary movements that spread over several successive touches.

Interpretation of results: When providing feedback for piano pedagogy, unjustified corrective feedback is problematic since this can lead to unnatural playing in order to satisfy the system. The student may exaggerate primary movements so that they are detected or tense up the muscles in order to avoid false positive detection of secondary movement. For discrete analysis of forearm rotation movements, primary movement detection rates of 99.87% were achieved while false positive detections occur with rates of up to 0.52%. In this case a single value of d can be used both to detect that the student executed a primary movement as supposed to and to detect that the student has avoided a primary movement as supposed to. For the other joints, where detection accuracy is less, it is advisable to modify the value of d according to the expectation of the feedback system: When the student is instructed to perform a primary movement, the value of d is reduced to reduce the probability of false negative detection. When the student is instructed to avoid a primary movement, the value of d is increased to reduce the probability of a false positive recognition.

The detection of primary movement in context of piano playing is particularly challenging as the primary arm movements are small while the secondary movements generated due to the mechanical interaction with the piano are relatively large. In other application areas, the separation between primary and secondary movement is usually clearer so that similar or better recognition rates can be expected.

Chapter 7.

Hand tracking

Contents

To use the analysis methods described in the previous chapter in a sensor-based feedback system, it is necessary to determine which hand has played a note (see Section 6.1). This chapter introduces and evaluates two methods and evaluates a third method that allow determining which hand has played a note:

1. The first method is based on MIDI data (Section 7.1).

2. The second method uses inertial sensors in addition to MIDI (Section 7.2).

3. The third method is based on computer vision (Section 7.3).

The third method was originally developed by Lefebvre-Albaret & Dalle for tracking hands for sign language recognition [116]. It is evaluated for tracking pianist hands here. The evaluation of all three methods is provided in Section 7.4. Section 7.5 compares the methods to related work and also compares the methods among each other.

Other application areas: Methods for hand tracking can be useful for other applications:

- **Hand-instrument mapping:** Electronic keyboards such as the Korg X5 [111] often allow the player to separate the claviature into two areas, one for the left hand and one for the right so that the player can play a different sound with each hand. To do so, each hand is confined to a fixed area, which contradicts normal piano practice. The methods proposed here can eliminate the need for this static boundary and enable a more natural playing experience.

- **Notation:** To notate a MIDI recording of a piano performance, it is necessary to perform hand assignment to assign the notes to the correct staff. Current notation software typically assigns notes to hands based on the note's position relative to a split point. The methods presented here can minimize the amount of post-editing that has to be performed by the user.

7.1. MIDI-based hand assignment

The methods for MIDI-based hand assignment and hand assignment based on sensor and MIDI data are closely related. Both methods are composed of a series of two steps. In the first step, a received note-on event is assigned to the left or right hand. In the second step, the note-on event is used to modify the estimated position of the hand.

7.1.1. Note assignment

Hand assignment of a note is done with two mechanisms: the identification of unique notes and the examination of the distances of the played note to the estimated hand positions. The method does not allow crossing over of the hands so that the left hand has to be located left of the right hand. It is possible to find simultaneously pressed keys that are located too far from each other to be played by one hand, a condition that will be called a *unique note*. As the hands are not allowed to cross over unique notes can be directly assigned to the left or right hand. Unique notes are identified as notes with an interval of more than an eleventh to the highest or lowest currently pressed key as most players cannot grasp such intervals. If a note is not an unique note, it is assigned to a hand based on the distance of the note to the estimated hand positions.

The positions of the hands are estimated with a Kalman filter for each hand. The received note-on events are handed over to the Kalman filter of the assigned hand.

7.1.2. Position estimation

For each hand, a Kalman filter [99, 178] is used to estimate the position of the hand. The state p of the filter is the position of the center of the hand. The position p is expressed in MIDI units. For example, let the center of the hand lie between the keys corresponding to MIDI pitch values of 60 and 61. Then the position p would be 60.5. The uncertainty of the position is expressed by the variance σ_p^2. The uncertainty of the position decreases when a measurement of hand position is obtained and increases otherwise.

A received note-on event is interpreted as an approximate measurement of hand position. The variance σ_m^2 expresses the uncertainty involved in the measurement. When a note-on message is received, the variance expressing the uncertainty in the position prior to incorporating the measurement $\sigma_p^2(t_2^-)$ is computed. Let t_1 be point in time when the last note was assigned to the Kalman filter and t_2 be the point in time when the new note was received. The uncertainty of the position before incorporating the new measurement $\sigma_p^2(t_2^-)$ is then updated based on the time difference between the two notes $t_2 - t_1$, a constant term σ_s^2, and the previous uncertainty after incorporating the measurement $\sigma_p^2(t_1^+)$.

$$\sigma_p^2(t_2^-) = \sigma_p^2(t_1^+) + (t_2 - t_1) \cdot \sigma_s^2 \tag{7.1}$$

The uncertainty of the position after incorporating the measurement $\sigma_p^2(t_2^+)$ is updated based on the uncertainty of the position before incorporating the measurement $\sigma_p^2(t_2^-)$ and the constant term σ_m^2 that expresses measurement uncertainty.

$$\sigma_p^2(t_2^+) = \sigma_p^2(t_2^-) - \frac{\sigma_p^2(t_2^-)}{\sigma_p^2(t_2^-) + \sigma_m^2}\sigma_p^2(t_2^-) \tag{7.2}$$

Let n be the pitch of the note received at t_2. Then the new position $p(t_2)$ is estimated based on the old position $p(t_1)$, the uncertainty of the position before incorporating the measurement $\sigma_p^2(t_2^-)$, and the pitch of the received note n.

$$p(t_2) = p(t_1) + \frac{\sigma_p^2(t_2^-)}{\sigma_p^2(t_2^-) + \sigma_m^2}(n - p(t_1)) \tag{7.3}$$

The values for σ_s^2 and σ_m^2 were empirically determined.

7.1.3. Discussion

This section illustrates the method with an example. Say, a user repeatedly plays two notes that are one octave apart with one hand. The first note is played after the hand has been inactive for some time. Therefore, the uncertainty of the hand position is high according to Equation 7.1. Because of the high uncertainty, the new measurement has great influence on the estimated hand position according to Equation 7.3 and the new estimated position will be near the pressed key. The uncertainty of the position reduces because of the new measurement according to Equation 7.2. Because the position uncertainty has been reduced, the next note, which is played one octave apart, receives less weight so that the new position is between the first and second note, slightly towards the second. After several touches, the position uncertainty levels off at a low value controlled by Equations 7.1 and 7.2 and execution speed. Therefore, new measurements do not drastically change the estimated position. The estimated hand position lies between the two alternating notes and only slightly oscillates when new measurements are made. If the user changes the position of the hand, the estimated position will adapt as older measurements loose influence over time according to Equation 7.1.

7.2. Hand tracking with sensors and MIDI

The method described in the previous section can be improved by using measurement of arm movement. This section details on the method based on inertial measurement and MIDI.

To re-position the hand, a player can use various movements of the arm and the body. Despite the many possibilities to move the hand to a given position, players usually reach a position with consistent body and arm posture. Therefore, the angle between the player's forearm and the keyboard can be interpreted as an indication for the position of the hand. The rate of change of this angle can be obtained from an inertial sensor attached to the wrist of the player. However, this measurement provides only information of position change. To obtain absolute hand position, the inertial measurement is combined with the MIDI through Kalman filtering.

Similar to the MIDI-based method, unique notes are assigned to the corresponding hand; non-unique notes are assigned to the hands based on the distances of the played note to the positions of the hands.

7.2.1. Arm movement measurement

To determine the rate of change of the angle between the forearm and the keyboard, which will be called the rate of sideways movement for simplicity, it is necessary to obtain the orientation of the sensor toward gravity. It would be possible to calculate pitch and roll angles directly from the accelerometer signal. However, the playing movements create additional sources of acceleration, which would adversely affect the accuracy. To improve the accuracy of the calculated pitch and roll angles, Kalman filtering is used to fuse accelerometer and gyroscope signals. Given the pitch and roll angle, the rate of sideways movement is calculated from the gyroscope signals.

7.2.2. Posture measurement

It is necessary to be able to convert a given angle between forearm and keyboard to a hand position (in MIDI pitch units) and vice versa. Because of different movement habits, the relation between playing position and angle has to be measured for each player individually. To this end, the player executes several touches with the same finger in a distance of, for example an octave, over the entire playing range of the keyboard. The change of the angle between two played notes is measured by summation of the rate of sideways movement. The measurement has to be performed for both hands. To convert from hand position to the angle between forearm and keyboard and vice versa, linear interpolation is used.

7.2.3. Signal fusion

For each arm, a Kalman filter is used to fuse MIDI and inertial measurement data. The state of the filter is the angle θ between the forearm and the keyboard.

When a new inertial measurement sample is received, the angle is updated. The new angle θ_{i+1} is computed based on the previous angle θ_i, the rate of sideways movement s, and the sample time dt:

$$\theta_{i+1} = \theta_i + s_i \cdot dt.$$

Crossing over of the hands is not supported and is avoided by setting s to zero if it would lead to a crossing over condition.

The uncertainty of the angle θ is expressed by the variance σ_θ^2. The uncertainty of the angle θ increases based on σ_s^2, which is the variance of the rate of sideways movement, and the sample time dt:

$$\sigma_{\theta,i+1}^2 = \sigma_{\theta,i}^2 + \sigma_s^2 \cdot dt.$$

When a note is assigned to the Kalman filter, the corresponding angle has to be calculated (see section 7.2.2). The measurement has an effect on the estimated angle and reduces the uncertainty of the angle. Let ϕ be the angle that corresponds to the pressed key that is assigned to the Kalman filter. The new estimate of the angle θ_{i+1} is calculated based on the previous angle θ_i, the previous uncertainty of the angle $\sigma_{\theta,i}^2$, and the angle ϕ:

$$\theta_{i+1} = \theta_i + \frac{\sigma_{\theta,i}^2}{\sigma_{\theta,i}^2 + \sigma_m^2}(\phi - \theta_i).$$

The uncertainty of the position is calculated based on the previous uncertainty and the measurement accuracy which is represented by the constant σ_m^2:

$$\sigma_{\theta,i+1}^2 = \sigma_{\theta,i}^2 - \frac{\sigma_{\theta,i}^2}{\sigma_{\theta,i}^2 + \sigma_m^2}\sigma_{\theta,i}^2.$$

The values for σ_s^2 and σ_m^2 were empirically determined.

7.3. Hand assignment with computer vision

The method discussed in this section was developed by Lefebvre-Albaret & Dalle for tracking hands for sign language recognition [116]. In the following its application to pianist hand tracking is discussed. The pixels belonging to the hands are detected by their color. Hand detection is performed with an annealed particle filter, where each hand is tracked by one cloud of particles. During the tracking process, the cloud pixels of each hand are alternatively subtracted from the skin detection map so that each cloud converges to a different hand. Hand positions are located at the centers of gravity of the particle clouds. Further details on the method can be found in [79, 116].

Hand assignment is performed by comparing the horizontal position of the played key in the video with the boundaries of the hands. The decision procedure takes into account whether the played key is inside the span of one hand, both hands, or outside both hands. If the key is inside the span of both hands, it is assigned to the hand where the key more inside the hand span. If the key is outside the span of both hands, it is assigned to hands based on distance. If one hand is outside the keyboard area, no notes will be assigned to it. To calculate the horizontal position of the played key, the procedure uses information about the keyboard position in the video, which is provided once by the user in a visual configuration dialogue.

7.4. Evaluation

To evaluate hand assignment accuracy, performances of different piano pieces were analyzed with our methods. To this end, video, MIDI, and inertial measurements that were performed with the MotionNet sensor system were recorded with one pianist playing different pieces.

Simple approaches to automatically evaluate the hand assignment results, for example by using score-following to match the obtained separation with a given correct separation, are problematic because of playing errors and differences because of ornamentation. Therefore, the results were manually examined. To this end it was necessary to present the result in a human-readably way. The text-based GUIDO format [89, 153] is used to create graphical musical scores for the left and right hand part. The human reader can then identify correct and wrong assignments.

The recorded pieces were the Sinfonias 1–5 by J. S. Bach (BWV 787–791) and the "Six Dances in Bulgarian Rhythm" (No. 148–153) from Bartok's Mikrokosmos vol. 6. Bartok's dances contain many instances where the hands overlap. Furthermore the dances contain frequent changes of hand position on the keyboard, which are performed very quick re-positioning movements. Also the hands are crossing over several times. Therefore, the dances are especially challenging for hand assignment.

The accuracies of the obtained hand assignments are shown in Table 7.1. For comparison with a baseline, the results of hand assignment with the split point method is included (split point is the Middle C). For all examined pieces, the proposed methods achieve better results than the split point method. The sensor-based method typically achieves better results than MIDI-based method. The camera-based method typically achieves better results than the MIDI-based and sensor-based methods.

The Bulgarian dance No. 152 shows a limitation of the methods. The hands often completely overlap in this piece, i.e., one hand is positioned above the other hand while both hands play notes in the same range. This is contrary to the assumptions of the methods, as they implicitly split the keyboard at a (time-variable) split-point. Therefore, the methods are not able to perform hand-assignment correctly if, e.g., the right hand plays a note that lies between two notes that are played with the left hand.

The sensor- and MIDI-based methods could be improved by using a hand model that filters out hand-note configurations that are impossible to grasp, which could be used instead of the simpler unique-note mechanism used here.

7.5. Discussion

MIDI-based methods: Kilian & Hoos proposed a method that finds a separation of a piece into different voices for notation [104]. Chords can occur in one voice. The

Table 7.1.: Hand assignment accuracy (indications in %)

Piece	Split	MIDI	Inertial	CV
Sinfonia 1	86.6	97.6	98.6	97.8
Sinfonia 2	86.4	94.3	97.0	98.6
Sinfonia 3	94.2	97.3	98.3	99.7
Sinfonia 4	90.9	97.5	98.6	98.9
Sinfonia 5	97.2	99.4	99.3	99.6
No. 148	81.8	88.7	91.2	94.2
No. 149	79.4	82.5	89.6	88.1
No. 150	81.9	86.4	83.6	90.8
No. 151	69.8	83.9	87.8	93.6
No. 152	65.2	66.6	68.4	70.6
No. 153	78.2	85.6	91.2	92.5

method allows the user to select the number of present voices. Therefore, it can be used to find a left hand and right hand part of a MIDI performance (see [104] for notated examples). To separate voices, the method by Kilan & Hoos splits the piece into a sequence of slices with overlapping notes and finds the voice separation by minimizing an elaborate cost function using a stochastic local search algorithm. This approach, while reasonable for notation, cannot be used for real-time hand assignment of a live performance because the slice of overlapping notes cannot be immediately determined when a note is received. Furthermore, the stochastic local search algorithm operates on the entire piece. Other voice separation methods [27, 98, 119] do not allow chords inside a voice and can therefore not be used for hand assignment.

Methods based on computer vision: To detect the two hands in the video, most of the studies make use of a skin color model. To be more robust to illumination changes, other color spaces than RGB, such as YUC or HSV, are often used for hand tracking. The hand color distribution can then be modeled as a histogram, a mixture of Gaussian, or any other parametric model [176]. Recent studies propose to combine the color information with a displacement information between two consecutive frames [70]. Hands are then identified by their motion and color. Edge detection is often used to refine the estimation of hand shape. After the hand pixels are detected, several algorithms can be applied for hand tracking (an overview is provided in [121]). Algorithms such as CamShift, CONDENSATION, etc. give very robust and accurate results as long as there is no hand occlusion. However, they often fail at labeling the right and left hand correctly after a big occlusion. However, overlaps and occlusions frequently occur in piano playing.

Gorodnichy & Yogeswaran developed a system for hand assignment that relies on visual tracking [60]. The system finds the position of the keyboard in the video and identifies the Middle C key. Background subtraction is used to find the hands in the image. Through the identification of cervices in the hand image, fingers are detected

Table 7.2.: Comparison of hand tracking methods

	Camera	Sensors	MIDI
Accuracy	1st	2nd	3rd
Comput. effort	High	Low	Low
Conditions	Contr. environment	No restr.	No restr.
Set-up effort	Some	No	No

although with correct labeling in only about half of the cases. The system annotates MIDI recordings with hand and finger labels.

Comparison of the proposed methods: The camera-based method provides the best accuracy, closely followed by the sensor-based and the MIDI-based method. When using it as part of a sensor-based feedback system, the sensor-based method has the advantage over the camera-based method that no additional hardware has to be set up since the sensors are already worn by the player to capture the piano playing movements. Furthermore, the lighting conditions do not have to be controlled. The MIDI-based method is best suited to support existing notation software to minimize the amount of post-processing that has to be done to assign the notes to the correct note system when notating a recording of a performance at the piano as it does not require any other data sources than the MIDI signal. The comparison is summarized in Table 7.2.

7.6. Summary

This chapter introduced and evaluated two methods and evaluated a third method that allow determining which hand has played a note:

1. The first method is based on MIDI data.
2. The second method uses inertial sensors in addition to MIDI.
3. The third method is based on computer vision.

The first two methods are based on a series of two steps. In the first step, a received note-on event is assigned to the left or right hand based on the distance of the current estimate of hand position. In the second step, the note-on event is used to modify the estimated position of the hand. The method using inertial sensors in addition to MIDI uses measurement of horizontal displacements to update hand position continuously. In both methods, hand position is estimated using one Kalman filter per hand. The third method, detects the hands by their color in video images using one particle filter per hand. During the tracking process, the cloud pixels of each hand are alternatively subtracted from the skin detection map so that each cloud converges to a different hand.

This method was originally developed by Lefebvre-Albaret & Dalle for tracking hands for sign language recognition [116] and is evaluated for tracking pianists hands here.

All three methods are effective to determine which hand has played a note. The best results are achieved with the computer-vision-based method, which is closely followed by the sensor-based and the MIDI-based method. The MIDI-based method is usable for improving existing notation software, which typically split a MIDI recording of a keyboard performance at the Middle C key. With the MIDI-based method a distinct improvement over this split approach is possible. For our sensor-based feedback, however, it is not sensible to use the MIDI-based method as the player already has to wear sensors. The main advantage of the sensor-based method is that it does not rely on a controlled environment (lighting, floor color, etc.), which is necessary when tracking hands visually.

Chapter 8.

Pedagogical applications

Contents

This chapter describes the pedagogical applications that were developed based on the analysis methods, the hand tracking methods, and the inertial sensors presented in the previous chapters. As the feedback relates to arm movements, their importance for piano pedagogy is discussed at the very beginning in Section 8.1. In Section 8.2, existing approaches for sensor-based feedback are presented. One approach is to examine the student's movement to determine whether it conforms to a desired target movement. Our system described in Section 8.3 adopts this approach. The system is based on a piano pedagogical movement notation. It checks whether the movement is executed as indicated by the movement notation using the results from Chapters 4 and 6 and provides acoustic feedback. A user study with students of a music university shows that potential users judge that this system is useful for learning technique. Another approach for sensor-based feedback is to prepare the sensor data and present it to the users so that they can examine and interpret the signal themselves. Our visualization application presented in Section 8.4 visualizes sensor data, MIDI data, and the musical score and replays the audio/video recording of the performance. To facilitate manual analysis, the application provides the possibility to synchronize two performances of the same piece by the student and the teacher. This makes it easy to spot differences where a closer examination may be beneficial. The expected amount of secondary movement is indicated in the sensor graphs to support the users during manual analysis.

8.1. Importance of arm movements

Arm movements play an important role in current piano pedagogy as the following historical overview shows, which is based on a description by Gerig [54]. The overview begins with a discussion of the early clavier methods that evolved for keyboard instruments before the modern piano was invented and continues to discuss major trends in piano pedagogy that have evolved since.

The early clavier methods were characterized by a passive arm and active fingers. Arm movement was used to change the horizontal position of the hand and for chord playing. This technique was appropriate for the harpsichord and the clavichord, which are predecessors to the piano. The harpsichord action is based on a plectrum, which is connected to the key with an upright jack. When the player presses a key, the plectrum plucks the string. The loudness of the generated sound is mainly determined by the action of the harpsichord and depends on force only to a small degree. Excessive force, however, results in usually undesirable percussive noises. Therefore, finger activity was preferred over the activity of the stronger arm. As the harpsichord action is very light compared to the action of the modern piano, the forces generated by the relatively weak fingers were sufficient [54, p. 9–34].

The piano action, however, is much heavier. Furthermore, percussive key sounds are less noticeable. Despite of this, keyboard technique remained nearly unchanged during the transition from the harpsichord to the piano. The so-called finger school had a culmination in the work of Carl Czerny (1791–1857). Czerny taught many celebrated pianists (among them were Franz Liszt, Theodor Leschetizky, and Theodor Kullak) and contributed technical exercises as well as writings about piano technique. Czerny's études have a place in the curriculum to the present day and are used for training finger dexterity and for learning musical patterns such as scales, arpeggios, etc. [54, p. 103–120].

Ludwig Deppe (1828–1890) was one of the first influential pedagogues to emphasize the role of the arm in piano playing. Deppe contributed only few written records about his method but his teachings were spread by his students. After Deppe's death, a multitude of books that emphasized the role of the arm were published [54, p. 229–270]. The most influential follower of that trend was Karl Breithaupt (1873–1945), whose name is connected to the school of weight technique. An important aspect of the Breithaupt's method was the use arm weight and muscle relaxation to execute touches. Breithaupt has been criticized for marginalizing the role of the fingers and to emphasize the role of weight for playing the piano overly [54, p. 329–359]. Today, many pedagogues acknowledge both the importance of finger and arm movements [54, p. 447–491]. As the russian pedagogue Leonid Nikolaev put it: "Nothing by fingers without arm, nothing by arm without fingers" [cited in 108, p. 33].

8.2. Approaches for sensor-based feedback

Data presentation systems: There are different ways to provide sensor-based feedback for instrument pedagogy. One approach is to present the sensor signals without substantial analysis to the user. The user then examines and interprets the sensor signals and learns about the playing movements. One example is Riley's system to visualize EMG and MIDI data [155] (see Section 3.2.1, p. 31). Such data presentation systems have the disadvantage that the interpretation of the sensor data can be difficult for the users and that the sheer amount of data can make the interpretation task tedious. Furthermore, there is evidence that providing plain information without interpretation is not helpful for motor learning [160, pp. 373–374]. Therefore, a teacher has to help the student to interpret the sensor data. As teachers are not experts in interpreting sensor signals, they have to invest considerable effort to learn the necessary skills before they can use such systems in their lessons. The main advantage of data presentation systems is that they provide the flexibility to examine various aspects of the performance while other approaches tend to be more limited in their scope. Section 8.4 discusses our data presentation system.

Movement conformity checking: Another approach for sensor-based feedback is to determine whether the student's movement conforms to a desired target movement. Two types of systems can be distinguished:

- Non-adaptable systems come with a fixed set of supported target movements that cannot be changed by the users.
- Adaptable systems can be changed by the users to support different target movements.

Grosshauser's & Hermann's exercise system for violinists [64, 65] (see Section 3.2.3, p. 39) is an example for a non-adaptable system. Here the target movements are defined by the code that generates the sonification or haptic feedback. Further examples include the systems by Peiper et al. [146], Rasamimanana et al. [152], and Young [189], where the system is trained by the developer to distinguish different bowing styles (see Section 3.2.2). An example for an adaptive system is Grosshauser's et al. system that provides a recognition engine based on Ordered Markov Models (OMMs) that is trained with correct executions of the target movement by the student under supervision of the teacher (see Section 3.2.2) [63].

Discussion: The main advantage of non-adaptable systems is that they are immediately usable and require minimal preparation effort. However, they are also limited to a fixed (and usually also small) set of supported movements. Adaptable systems provide flexibility, which enables the teacher to create individual exercises. However, this flexibility is traded with preparation effort expended by the users to train the system. To minimize this preparation effort, it is sensible that the users provide only a single sample of the movement as suggested by Grosshauser & Hermann [63]. The decision

Table 8.1.: Comparison between adaptable and non-adaptable systems

System type	Flexibility	Preparation effort
Non-adaptable	Low flexibility	Low effort
Adaptable	High flexibility	High effort

whether the movement is executed correctly is then based the similarity between the new movement and the recorded one [63]. Since only a single sample is used for training, the recognition does not generalize well. To support feedback for the same movement in different conditions, e. g., playing with a different tempo, playing louder or softer, playing different pitches, etc., it is often necessary to provide the system with a sample of the movement in that specific condition. Furthermore, since students have individual movement characteristics, the system has to be trained individually for each student. Table 8.1 summarizes the discussion on adaptable and non-adaptable systems. Non-adaptable systems are inflexible but can be used by the users out-of-the box. Adaptable systems are highly flexible. This flexibility however is traded with high preparation effort.

Symbolical model: Ideally, a feedback system provides high flexibility with little preparation effort. Our approach towards this goal is to use a symbolical model of the target movement. This symbolical model of the movement is provided by the teacher using a special movement notation. The movement notation system has to be sufficiently abstract so that the teacher can notate the exercise with little preparation effort. Furthermore, the movement notation system has to be sufficiently expressive so that high flexibility can be reached.

8.3. Sonification

Instead of developing a new movement notation, we surveyed piano pedagogy literature to identify existing movement notations. We selected one particular notation and used the results presented in Chapters 4 and 6 to realize a feedback system that provides acoustical feedback, i. e., sonifies, whether the movements were executed as indicated by the notation.

8.3.1. Movement notations

The main movement notations that can be found in piano pedagogy literature can be grouped into four types, which we call the “posture sketch”, the “trajectory sketch”, the “movement signal notation”, and the “augmented score”.

Posture sketch: The posture sketch is a movement notation where a single or several depictions of the player's posture are used to provide the reader with an impression of the movement. Examples are drawings of the body, stick-figures, or photographs (see Figure 8.1). If a single depiction of a posture is used, additional markings such as arrows are often used to indicate the movement. Movement sketches have been used by various piano pedagogues including Matthay [123], Breithaupt [19], E. J. Bach [6], Gat [52], and S. Bernstein [10].

Trajectory sketch: A trajectory sketch is a notation where the two-dimensional projection of the moving body is drawn (see Figure 8.2). Time passes as one follows the trajectory, which contradicts the practice in traditional music scores where time passes from left to right. Therefore, the trajectory is often annotated to make the connection to the music score. Trajectory sketches have been used by various piano pedagogues including Breithaupt [18], Varró [174], E. J. Bach [6], and Marek [122].

Movement signal notation: Movement signal notation expresses aspects of the movement as a graph of an one-dimensional function over time (see Figure 8.3). E. g., one can notate the vertical movements of the wrist. Movement signal notation has the advantage that it combines well with a musical score as time passes from left to right in movement signal notation just as in musical scores. Movement signal notation has been used by various piano pedagogues including Breithaupt [18], E. J. Bach [6] and Kochevitsky [108].

Augmented score notation: Augmented score notation is based on a traditional music score that is augmented with additional symbols to describe the movement to perform the score (see Figure 8.4 and [6]). Augmented score notation has been used by various pedagogues including Breithaupt [20], E. J. Bach [6], Bernstein [9–11], and Marek [122].

Discussion: To be usable as a basis for sensor-based feedback, a notation system has to be compatible with music scores so that exercises can be notated easily. This excludes posture sketches and trajectory sketches. Posture sketches consume too much space so that it is impractical to use them in a music score. Trajectory sketches are not compatible as the passing of time in a trajectory sketch contradicts the passing of time from left to right in a music score. Movement signal notation combines well with traditional music scores. However, the continuous line implies an unrealistic level of detail. Of course, it is assumed that the reader does not take the graph entirely literally. To use movement this kind of notation, the feedback system would have to interpret the graph and understand the underlying idea of the movement behind the concrete representation just as a human reader would. Since this is non-trivial, we chose not to use this notation as basis for the sensor-based feedback system. Augmented score notation is compatible with traditional music scores and is unambiguous with the respect to the level of detail that a symbol provides. Therefore, augmented score notation was chosen as basis for our sonification system. Table 8.2 summarizes this discussion.

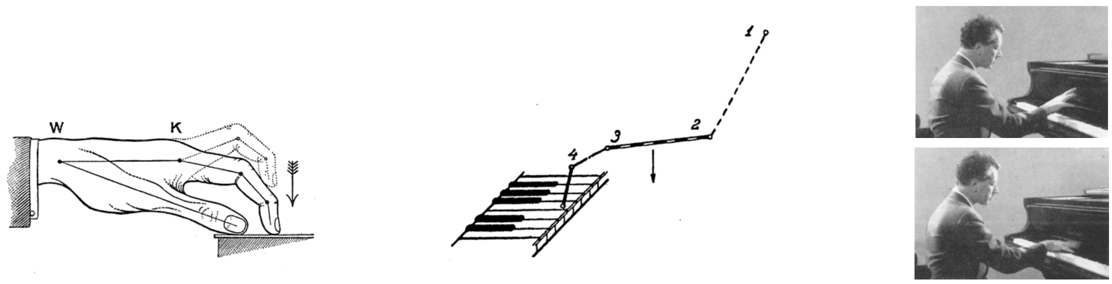


Figure 8.1.: Movement sketches: drawing (left) [123], stick-figure (center) [174], and photographs (right) [52]

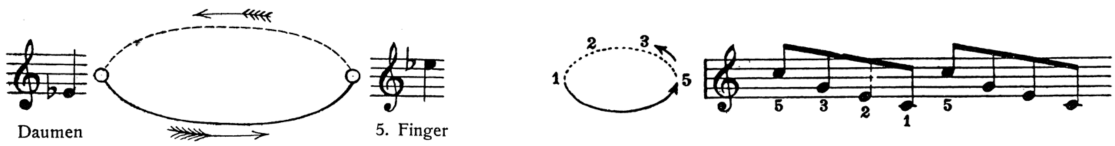


Figure 8.2.: Trajectory sketches: Breithaupt indicates the wrist trajectory to execute Chopin's Étude Op. 25 No. 1 by framing the trajectory with the lowermost and uppermost note (left) [18], Varró makes the connection to the musical score by marking the fingering on the wrist trajectory (right) [174].

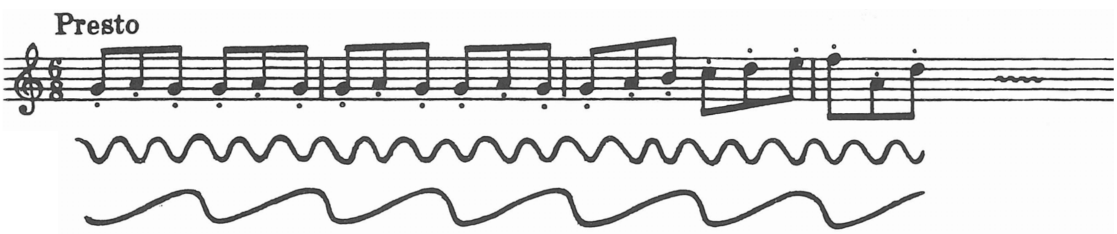


Figure 8.3.: Movement signal by Kochevitsky indicating wrist and upper arm movement [108]

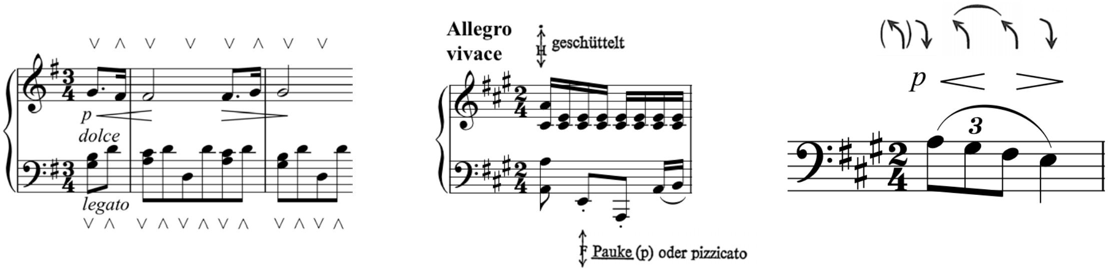



Figure 8.4.: Augmented scores by Breithaupt (left) [20], Marek (center) [122], and Bernstein (right) [10]

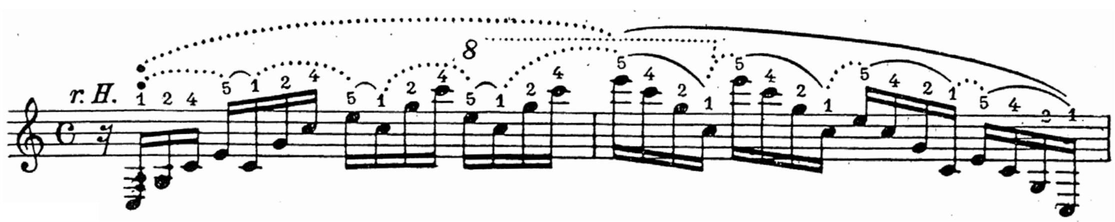



Figure 8.5.: E. Bach's augmented score of Chopin's Étude Op. 10 No. 1 [6]: The dotted and drawn-through slurs indicate the direction of forearm rotation. Bach's systems is recursive in the sense that it allows to expressing that an overall rotation in one direction is composed of smaller rotation movements, which may also include rotations in the opposing direction.

Table 8.2.: Comparison between different notation systems

Notation system	Compatible	Unambiguous
Posture sketch	No	**Yes**
Trajectory sketch	No	No
Movement signal notation	**Yes**	No
Augmented score notation	**Yes**	**Yes**

8.3.2. Realization

Augmented scores have been used by various pedagogues. It was decided to realize sensor-based feedback for the movement notation developed by the piano pedagogue S. Bernstein [9–11]. This has the advantage that existing exercise books [9, 11] can be supported with sensor-based feedback.

Touch movements: Bernstein's movement notation is based on two elementary movements: the vertical movement of the wrist and the rotation of the forearm. The vertical movement of the wrist is represented by the symbols "$\uparrow$" and "$\downarrow$". To move the wrist upwards, the upper arm moves forward (extension of the upper arm) while the hand is flexed in the wrist so that the fingers remain in contact with the keys. To move the wrist downwards, the upper arm moves backwards (flexion of the upper arm) while the hand is extended in the wrist. Forearm rotation is notated with the symbols L and R, which indicate a counterclockwise respectively clockwise rotation. The combination of the two elementary movements are indicated with the symbols "$\curvearrowright$", "$\curvearrowleft$", "⤵", and "⤶", which show the direction of vertical wrist movement and the direction of forearm rotation. Each of these symbols refers to a single note [10, 11]. To check whether the indicated movement was executed, it is necessary to determine whether primary movements in the correct direction have occurred in the said joints. By using the method of discrete analysis of single touches (as discussed in Section 6.2), this is straightforward to check.

Spread movements: Arm movement can be spread out over several successive notes. To notate these movements, Bernstein groups several instances of the same symbol under a slur:

- Vertical wrist movement is indicated with the symbols "$\overparen{\uparrow\uparrow\uparrow}$" and "$\underparen{\downarrow\downarrow\downarrow}$".
- Forearm rotation is indicated with the symbols "$\overparen{RRR}$" and "$\overparen{LLL}$".
- The combination of vertical wrist movement and forearm rotation is indicated with the symbols "$\overparen{\curvearrowright\curvearrowright\curvearrowright}$", "$\underparen{⤶⤶⤶}$", "$\overparen{\curvearrowleft\curvearrowleft\curvearrowleft}$", and "$\underparen{⤵⤵⤵}$" [10].

To check whether the indicated movement was executed, it is necessary to decide whether primary movement has occurred in the correct direction in the said joints over the indicated time interval. This is done with the method for serial analysis of successive touches presented in Section 6.4. In order to use serial analysis the beginning t^- and the end t^+ of the analysis interval have to be defined. The end of the analysis interval t^+ is defined as the note-onset of the last note to be played with the indicated movement. The beginning of the analysis interval t^- is the point in time 0.1 s before the note-onset of the first note, which is equal to the beginning of the discrete analysis interval of the first touch.

Preparation movements: Before a note is played, a preparation movement can occur. This preparation movement is often a movement in opposite direction of the movement that occurs when the note is played. Bernstein uses two notation variants to indicate preparation movements. On the one hand, he sets one of the movement signs "↑", "↓", "L", "R", "↷", "↶", "⤵", or "⤶" in parentheses. On the other hand, he groups the preparation movement with the following opposite movement under a slur: "$\overparen{\uparrow\downarrow}$","$\overparen{LR}$", "$\overparen{RL}$", "$\overparen{↷⤶}$", '$\overparen{⤶↷}$", "$\overparen{↶⤵}$", and "$\overparen{⤵↶}$" [10]. Detecting primary preparation movements is relatively easy since there is no mechanical interaction that leads to time-varying key reaction forces: Either there is no contact between finger and key or the key is held down, which leads to a constant, i. e., non time-varying key reaction force. In consequence only minimal secondary movement occurs. Therefore primary movement can be detected using thresholding with fixed boundaries as discussed in the following.

To detect whether there is primary preparation movement in the i-th degree of freedom of the arm in positive direction F_i^+, which is defined as

$$F_i^+(t) = \begin{cases} F_i(t) & \text{if } F_i(t) > 0, \\ 0 & \text{else,} \end{cases}$$

where $F_i(t)$ is the angular rate in the i-th degree of freedom of the arm. A primary movement is detected if

$$\int_{t_b}^{t_1 - 0.1\,s} F_i^+(t)dt > c,$$

where c is a fixed constant, t_1 is the onset time of the note that was prepared with the preparation movement and

$$t_b = \max\left(t_0 + 0.2\,s, t_1 - 0.5\,s\right),$$

where t_0 is the onset time of the previous note. Similarly a primary preparation movement in negative direction is recognized if

$$\int_{t_b}^{t_1 - 0.1\,s} F_i^-(t)dt > d,$$

where

$$F_i^-(t) = \begin{cases} -F_i(t) & \text{if } F_i(t) < 0, \\ 0 & \text{else.} \end{cases}$$

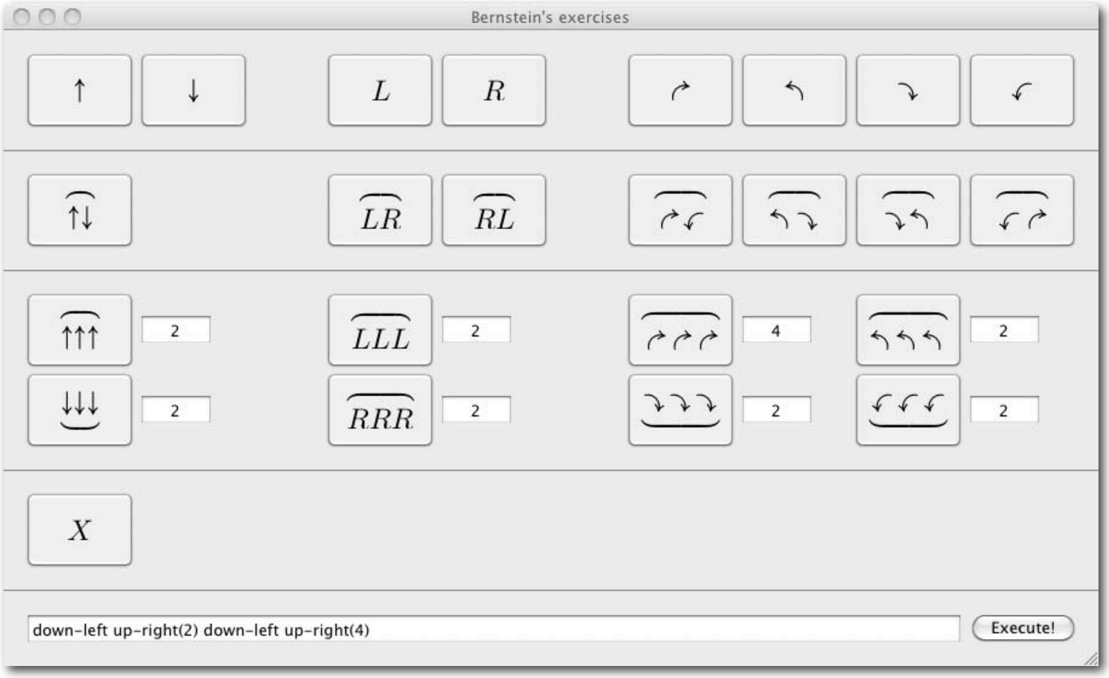


Figure 8.6.: A graphical user interface to define exercises in the style of Bernstein

Graphical user interface: Ideally, a plugin in existing music notation programs such as Sibelius or Finale would be used to notate exercises in the style of Bernstein. Since this would introduce considerable development effort while providing no scientific value, a simple graphical user interface to define exercises is used instead (see Figure 8.6). By clicking on the symbols, which are organized in four rows, one defines a sequence of movement symbols. The first row contains symbols that relate to a single touch. The second row contains symbols that indicate a touch movement that is accompanied by a preparation movement. The third row contains symbols that indicate a movement that is spread over several notes. The number of notes over that the movement is spread are defined in the text field next to the symbol. The forth row contains the single symbol "*X*", which is used by the user to indicate that the note may be played with any movement.

Feedback: When a movement is executed correctly by the user, the system generates a MIDI note-on message with the same pitch, which is routed to a software or hardware synthesizer. The system generates a MIDI note-off message when the key is released. By routing MIDI commands that relate to the pedal to the synthesizer, the player can use the pedal to prolong the synthesized notes. The normal piano sound is always audible to the user. The user hears the sound of the piano plus the sound from the synthesizer, when the indicated touch movement is executed correctly. If the movement is executed correctly, the last note of the sequence is doubled.

Coverage: Table 8.3 summarizes Bernstein's movement notation system. Symbols 1 through 6 indicate touch movements. Symbols 9 through 11 indicate touch movements that are prepared by an opposite preparation movement. Symbols 12 through 14 indicate movements that are spread over several notes [10]. The symbols 1 through 14 are supported by the application while the symbols 15 and 17 are not:

- Symbol 15 "P" indicates to place the fingers on the key [10]. Since no measurement of fingertip position is available from the used sensors, the application cannot check this.
- The symbols 16 ("→", "←") and 17 (') indicate horizontal movement of the hand along the keyboard [10]. Being able to play the correct notes shows that the necessary lateral movements are performed by the user, making it is unnecessary to analyze the sensor signal to check this (although this would be possible).

8.3.3. User study

Mobile prototype: A user study was performed with piano students of the Frankfurt University of Music and Performance Arts. The goal of the user study was to collect feedback on the applicability of the system for piano pedagogy. To minimize the student's time effort to participate in the study it was decided to perform the user study at the university of music in Frankfurt. Therefore, a mobile version of the system was necessary. However, the experimental results in Chapter 6 were performed using a Kawai K–15 ATX acoustic piano with MIDI interface. Therefore it was necessary to use a portable digital piano: We chose to use a Casio CDP-100 digital piano. However, using another instrument made it necessary to retrain the system since the mechanical properties of the two actions differ so that different amounts of secondary movement may result from identical key speeds. Furthermore, the mapping from MIDI velocity to key speed cannot be expected to be the same for two different instruments. As shown in Section 6.3.3, only relatively few samples are necessary for training. Therefore, the system was trained with 360 samples ($360 = 2 \cdot 180 = 2 \cdot 3 \cdot 6 \cdot 5 \cdot 2$), which corresponds to two samples per combination of different loudness levels (3), touch types (6), fingers (5), and direct vs. indirect touch (2). A stable keyboard stand was used to to make sure that the keyboard itself does not move, which is important to achieve best possible recognition accuracy. The values of d, which control the sensitivity of primary movement detection (see Equation 6.2 on p. 86) were empirically chosen so that the system provides a good user experience weighing detection sensitivity against false positive detection of primary movement. The value $d = 1$ was chosen to recognize primary hand extension/flexion in the wrist and arm extension/flexion in the shoulder joint. To recognize primary forearm rotation movement, a value of $d = 3$ was used.

Preparatory lecture: In preparation of the user study, the students participated in a preparatory lecture. The actual user study had to be restricted to a subset of the supported movements because of the students' time constraints. Therefore, it was necessary

Table 8.3.: Bernstein's movement notation [10]

	Movement	Sign
1	Wrist up	$\uparrow$
2	Wrist down	$\downarrow$
3	Rotate right	R
4	Rotate left	L
7	Wrist up, rotate right	$\Rsh$
8	Wrist up, rotate left	$\Lsh$
5	Wrist down, rotate right	$\curvearrowright$
6	Wrist down, rotate left	$\curvearrowleft$
9	Upper arm roll	$\overbrace{\uparrow\downarrow}$
10	Double rotations	$\overbrace{LR}$ $\overbrace{RL}$
11	Double rotations and upper arm rolls	$\overbrace{\Rsh\curvearrowleft}$ $\overbrace{\curvearrowleft\Rsh}$ $\overbrace{\Lsh\curvearrowright}$ $\overbrace{\curvearrowright\Lsh}$
12	Continuous upper arm movement	$\overbrace{\uparrow\uparrow\uparrow}$ $\underbrace{\downarrow\downarrow\downarrow}$
13	Continuous rotation	$\overbrace{RRR}$ $\overbrace{LLL}$
14	Continuous upper arm movement and rotation	$\overbrace{\Rsh\Rsh\Rsh}$ $\underbrace{\curvearrowleft\curvearrowleft\curvearrowleft}$ $\overbrace{\Lsh\Lsh\Lsh}$ $\underbrace{\curvearrowright\curvearrowright\curvearrowright}$
15	Fingers are placed on the keys	P
16	Horizontal movement	$\rightarrow$ $\leftarrow$
17	Jump	'

to provide a complete overview of the system and Bernstein's movement notation beforehand. The preparatory lecture featured a brief pedagogical background of Bernstein's "school of movement" and the pedagogical motivation of our feedback system. The main part of the lecture consisted of theoretical explanations of the notation symbols and also practical demonstration of a subset of the symbols with the aid of videos. Each video showed the execution of the movement together with the feedback of the system. In that way the students became acquainted with both the notation and the feedback system. Figure 8.7 shows the eight examples that were demonstrated to the students on video. The lecture lasted for approximately one hour.

Execution of the study: Eight students were recruited to participate in the lecture and in the subsequent user study. The students were recruited by a piano professor and a student helper. The user study took place two days after the lecture. Five students were pursuing a degree in piano pedagogy (in German: "Instrumental- und Gesangspädagogik") and two students were studying towards an artist diploma (in German: "Künstlerische Ausbildung"). One student that had participated in the lecture was not able to come to the user study so that a total of seven students participated in the user study. We met individually with each participating student. Each participant tried out the system playing the examples 1 through 4 shown in Figure 8.7. When an example was mastered by the student or when only little further improvement seemed possible in the available time, the student was instructed to move along to the next example. The system was set up to recognize the movement covered in the exercise and provided acoustic feedback in form of a synthesizer sound that doubled the played note if the system movement was recognized. Before the student began to play an exercise, we reminded the student of the movement by explaining and demonstrating it once again. If the student had difficulties executing the required movement, further hints were provided. The hints were one of the following:

- **Correction:** "You did [description of the movement performed by the student] but the system expected [description of the target movement]."
- **Preparation movement:** "If you want to move in this direction when you play this note, you should move in the opposite direction when you play the other notes."

The students were typically able to execute some movements without effort while other movements took some practice. Which movements were difficult varied between the students. After a student had performed the exercises with the system, a questionnaire was filled out. A session with one student typically lasted for approximately 30 minutes.

General questions: The questionnaire consisted of a series of statements that were rated with a five-point Likert scale. The questionnaire consisted of two parts: a first part with general questions and a second part with questions about the system. The general questions served to collect the students' opinions on some of the pedagogical

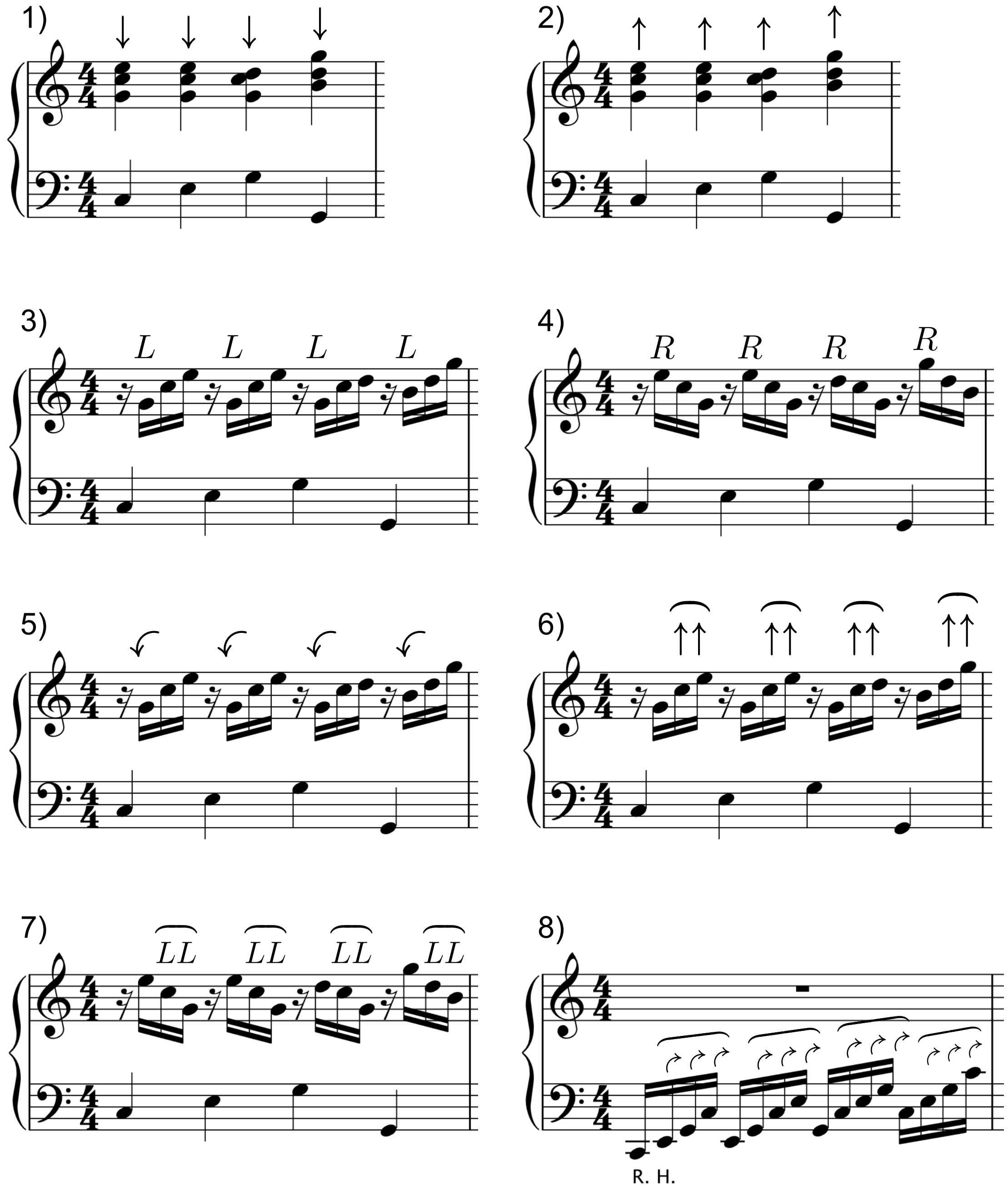



Figure 8.7.: The material that was shown to the students on video. Examples 1 through 7 are adapted from the Prelude C-major BWV 924 by J. S. Bach. Examples 1 through 7 lack the embellishments in the left hand that are present in the original. They were played up to bar 9. Example 8 shows a C-major arpeggio, which was played up to the c$''''$. In the user study, the students were instructed to play examples 1 through 4.

Table 8.4.: The results of the general part of the questionnaire.

Statement	M	SD
Arm movements are important for a good technique.	5.0	0.0
It is worthwhile to occupy oneself with the playing movements of the arm.	4.9	0.4
Bernstein's school of movement can help to acquire a better technique.	3.6	1.0
Bernstein's school of movement can help to achieve better musical expression.	2.9	0.9
I would consider to use Bernstein's movement notation when teaching piano students.	3.3	1.1
I would consider to occupy myself with Bernstein's "school of movement" to improve my own play.	3.3	1.4

foundations of the system. The students unanimously agreed that arm movements are important for a good technique (M = 5, SD = 0). They also agreed that it is worthwhile to occupy oneself consciously with the movements of the arm (M = 4.9, SD = 0.4). The students were more reserved towards Bernstein's school of movement. Yet, the question whether Bernstein's school of movement can help to improve technique was rated positive (M = 3.6, SD = 1.0). The question whether Bernstein's school of movement can help to improve the musical expression was rated slightly negative (M = 2.9, SD = 0.9). Slightly positive scores were also given by the students on the question whether they would consider to use Bernstein's movement notation when teaching piano students (M = 3.3, SD = 1.1) and using Bernstein's school of movement to improve their own playing (M = 3.3, SD = 1.4). However, as already evident in the high standard deviations, the students did not agree whether they would want to use Bernstein's school of movement for their own teaching and practicing. The results are summarized in Table 8.4.

System-specific questions: The second part of the questionnaire contained question about the system. The students agreed that the system recognizes their movements with high accuracy (M = 4.9, SD = 0.4). The system was experienced to be fun to use (M = 4.7, SD = 0.5) and to increase the motivation to occupy oneself with playing movements (M = 4.9, SD = 0.4). The students were positive that one can learn something useful

Table 8.5.: The results of the part relating to the system

Statement	M	SD
The system recognizes my movements accurately	4.9	0.4
Using the system is fun	4.7	0.5
The system increases the motivation to occupy oneself with playing movements	4.9	0.4
One can learn something useful about one's playing movements with the system	4.6	0.8
One can learn something useful to improve technique with the system.	4.3	0.9
One can learn something useful for improving one's musical abilities with the system.	2.6	1.2
I would want to use the system to improve my knowledge on playing movements	4.0	0.8
I would want to use the system when teaching piano students	3.7	0.9

about one's playing movements with the system (M = 4.6, SD = 0.8) but rather for improving technique (M = 4.3, SD = 1.0) than one's musical abilities (M = 2.6, SD = 1.3). Finally, the participants were asked whether they would want to use the system for themselves and whether they would want to use the system for teaching their piano students. These questions weigh the advantages of using the system for the training of technique with its disadvantages, e. g., having to put on the wearable sensors and having to set up and operate a computer system. Nevertheless, the participants tended to be positive that they would want to use the system for themselves (M = 4.0, SD = 0.8) and to use it to teach their piano students (M = 3.7, SD = 0.9). The results are summarized in Table 8.5.

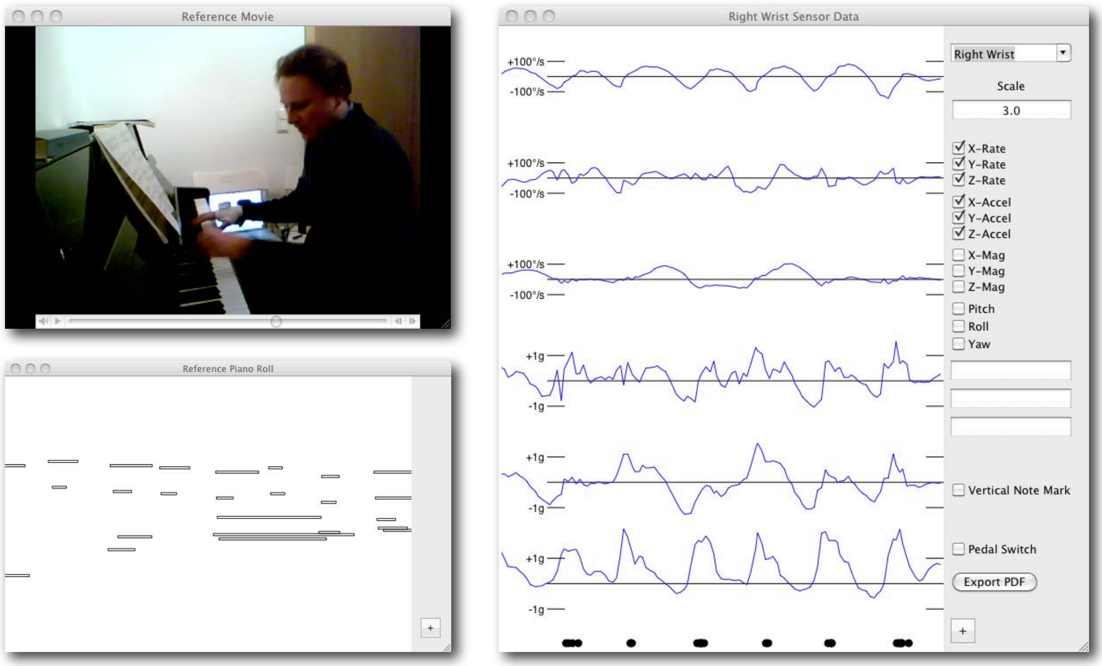


Figure 8.8.: Sensor data, piano roll, and video are shown

8.4. Visualization

Visualization: The second pedagogical application (see Figure 8.8) visualizes sensor data, MIDI data, and the musical score (see Figure 8.9) and replays the audio and video recording of the performance. The user can choose between visualizing the raw sensor data or the joint angular rates calculated for each of the seven main degrees of freedom of the arm. A piano roll representation of the performance provides the users with an exact visual representation of the performance. The users can examine the onsets and offsets of the notes to examine rhythm and articulation of the performance in-depth. Audio and video have two functions. On the one hand, they provide an additional modality to examine the performance. On the other hand, they provide the users with a sense of orientation in the piece. Likewise, the rendering of the score (see Figure 8.9) provides better orientation to the users. A horizontal line marks the current position. To use the animated score, the users have to provide images of the score to the system, e. g., by scanning the paper score. To synchronize the moving horizontal line with the video, audio, and sensor data, the users have to provide a beat track and mark beats in the score images using a special editor. An in-depth description of the animated score and the editing software is provided in [13, 71] and will not be repeated here.

Performance synchronization: Analyzing sensor data by hand is challenging. To simplify manual analysis, the visualization system allows synchronizing two performances

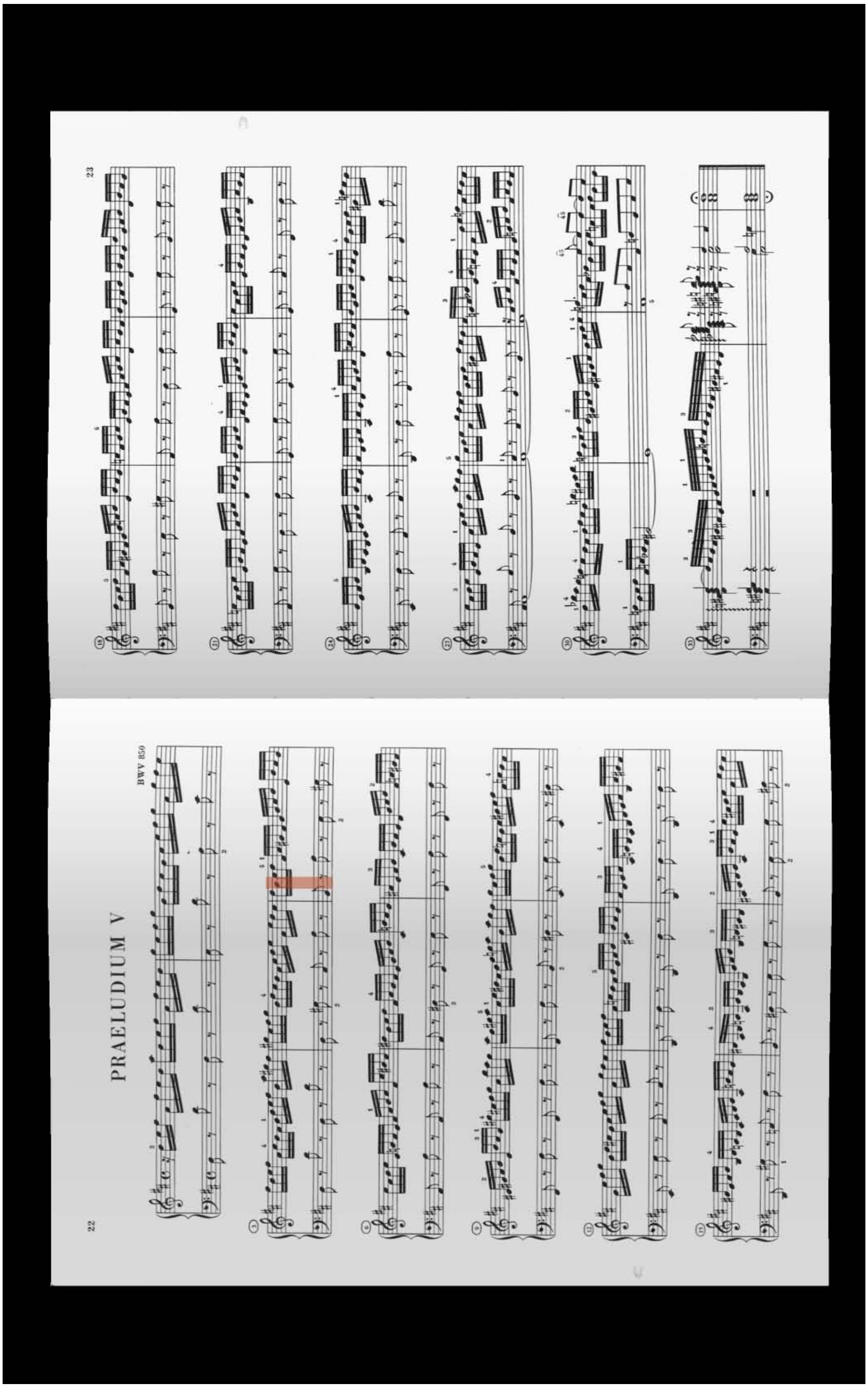


Figure 8.9.: An animated music score provides orientation to the users.

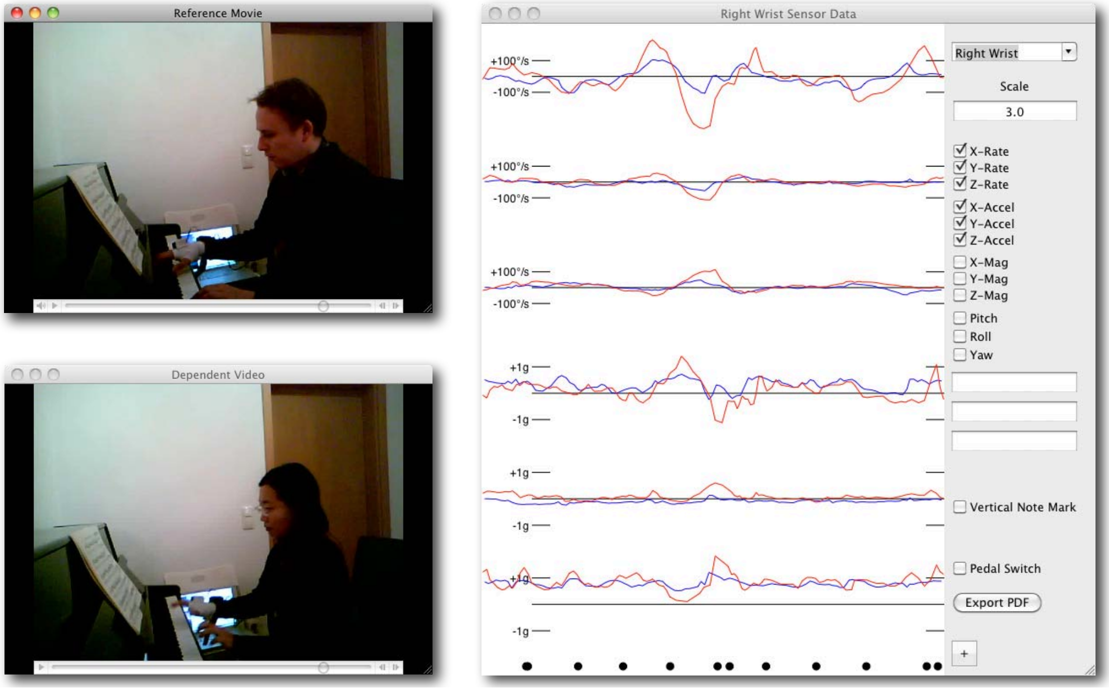


Figure 8.10.: Comparison of two performances

of the same piece performed by the teacher and a the student. This makes it easy for the users to spot differences where a closer examination may be beneficial. The visualization system distinguishes the reference performance, which remains unchanged, and the dependent performance, which is time-stretched to fit to the reference performance. Score following is used to identify corresponding notes in the two performances. The video and the sensor signal of the dependent performance is time-stretched. To make it easy to spot differences, the sensor signals of the two performances are placed over each other in a single graph and drawn with different colors (see Figure 8.10).

Secondary movement analysis: Manual analysis of sensor data is complicated by the presence of secondary movement. Users can be in doubt whether a deflection in the sensor signal is due to a primary movement. Since secondary movement is not directly conscious controllable, users are typically not interested to examine it closely. To provide a hint to the users to understand what deflections of the sensor signal can be ignored, the expected amount of secondary movement, which is determined by discrete analysis, is visualized for each touch in the graphs showing the angular velocities in the arm joints (see Figure 8.11). When the MIDI interface reports a note onset, the analysis is triggered. The amount of movement that occurs in the analysis interval, i. e., F_i, is determined and visualized as a bar. The amount of secondary movement that can be expected is indicated by three horizontal line that are drawn over the bar: the middle

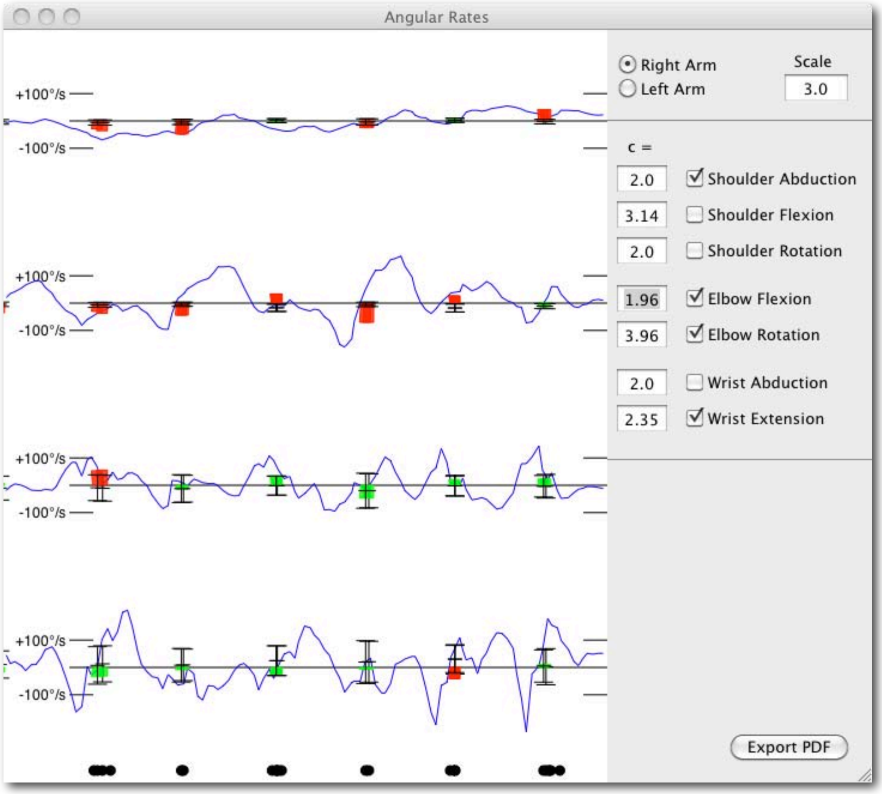


Figure 8.11.: Visualization of the analysis results

line indicates the mean μ_i, the other two indicate $\mu_i \pm d \cdot \sigma_i$ of the secondary movement, where d is the optimal value as determined in Section 6.3.2 by default. If the bar exceeds one of the outer lines, i. e., if the amount of secondary movement exceeds $\mu_i \pm d \cdot \sigma_i$, the bar is filled with red color to indicate that a primary movement occurred. Otherwise, the bar is filled with green color to indicate that only secondary movement occurred.

8.5. Summary

Starting with Deppe and his followers in the late 19th century, the importance of arm movements for piano technique was recognized [54, p. 229–270]. Today many pedagogues acknowledge that both arm and finger movements are important [54, p. 447–491]. This chapter presented applications of the movement analysis methods, hand tracking methods, and inertial sensors presented in the previous chapters. The applications provide feedback on arm movements.

There are different approaches to sensor-based feedback. One approach is to determine whether the student's movements conform to a desired target movement. This approach was adopted in an application that sonifies the user's arm movements. To make the connection with existing piano pedagogy practice, movement notations used in the pedagogical literature were examined. We chose to support augmented score notation, which is a notation where additional symbols are added to a traditional music score to indicate the movement. Using the analysis method introduced in Chapters 4 and 6, a particular notation invented by the piano pedagogue S. Bernstein was realized. A user study with students of a music university shows that potential users think that the system is useful for learning technique.

Another approach for sensor-based feedback is to provide the sensor data to the users and let them analyze and interpret the signal. Our second pedagogical application adopts this approach. The visualization application visualizes sensor data, MIDI data, and the musical score and replays the audio/video recording of the performance. Based on the sensor data, the teacher and the student can examine the movements. MIDI data is displayed as a piano roll, which allows the users to examine rhythm and articulation visually. The audio/video recording provides additional modalities to examine the performance. A rendering of the score, which includes a marking that indicates the current position, provides orientation to the users. To facilitate manual analysis, the performance of the same piece played by different persons can be synchronized so that video and sensor data can be compared. This makes it easy to spot differences where a further examination may be worthwhile. To help the users to cope with secondary movement, the expected amount of secondary movement is marked in the sensor graphs.

Chapter 9.

Summary and future work

Secondary movement in one part of the body results from primary movement, i. e., goal-directed, consciously controllable movement, in other parts of the body and from the mechanical interaction with the environment. When a primary movement is performed, time-varying joint torques are generated throughout the body. Biomechanical analysis differentiates between inter-segmental interaction torques that originate from movements of proximal or distal limbs and reaction torques that are generated by an reaction force, which is transmitted through the chain of limbs to various joints of the body [49]. The body can counteract these time-varying torques by contracting the appropriate muscles or by tensing up agonist and antagonist muscles to stabilize a joint in preparation. However, the time-varying torques cannot be completely compensated and thus lead to small secondary movements. Furthermore, many muscles have an effect on more than one joint [23] making it difficult to move a limb in total isolation is a further source of secondary movement.

This thesis introduced methods to distinguish between primary and secondary movement in sensor signals of human movement. The methods were applied to the analysis of piano playing movements and evaluated in this context. Piano pedagogical applications were developed based on the movement analysis methods. The analysis methods are generally useful to analyze a certain type of motor tasks, namely motor tasks that have a significant amount of secondary movement. Furthermore, the analysis methods can be used as a preprocessing step to improve current gesture recognition techniques by removing secondary movement from the sensor signal.

9.1. Summary

In the following the main contributions of this thesis are summarized.

Movement analysis: This thesis introduced two methods to distinguish between primary and secondary movement: discrete and serial analysis. Discrete analysis is based on a data set of samples of secondary movements. Based on that data set and measurements of factors that influence secondary movement, the amount of secondary movement can be estimated. By comparing the measurement of a new movement with the estimation of secondary movement, discrete analysis decides whether a primary movement occurred based on a parameter to control recognition sensitivity.

While discrete analysis is limited to analyzing fixed and usually short time intervals, serial analysis allows combining several successive discrete analyses to determine whether a primary movement has occurred during a larger time interval. This is possible without the need of additional data collection or training. This advantage, however, is traded with a reduced accuracy due to simplifying assumptions. Serial analysis combines the estimations of secondary movement by the discrete analyses contained in the time interval to obtain an overall estimation of the secondary movement. Then the movement measurement is compared to this estimation to decide whether a primary movement occurred based on a parameter to control recognition sensitivity.

Discrete and serial analysis were applied to analyze pianist arm movements, which is a field where it is particularly difficult to separate primary and secondary movements. Discrete analysis was used to detect primary arm movement when playing a single note. The amount of secondary movement was estimated based on an estimation of key reaction force from MIDI data. Furthermore, the estimation of secondary movement in one joint depended on the movements registered in the other joints of the arm. The estimation of secondary movement was learned on the basis of an extensive data set using maximum likelihood estimation. Serial analysis was used to detect primary movement spread over several successive notes. An evaluation shows that both discrete and serial analysis are able to distinguish primary from secondary arm movements. The accuracy of the recognition is in general high but depends on the examined joint. Primary forearm rotation movements are detected in 99.87% of the cases. False positive detections of primary rotation movements range from 0 to 0.52%. Primary wrist and elbow flexion/extension movement is detected with the least accuracy: Primary movements are detected with a rate of 92.36 in the wrist respectively 91.16% in the elbow. False positive detections are also more common in the wrist and the elbow ranging from 0.94 to 7.83% (wrist) and from 2.52 and 3.69% (elbow). The reason for the lower detection accuracy for elbow and wrist movements is that these movements are particularly small in comparison with the secondary movements that occur in these joints.

To analyze a series of successive touches, discrete analysis is not applicable since this would require collecting a prohibitively large data set as the possible variations increase exponentially with the number of notes. Therefore, serial analysis was used for this. An evaluation shows that serial analysis is effective. The accuracy, however, is reduced in comparison to discrete analysis: The detection rate of primary forearm rotation movement when performing serial analysis is 96.7% with a rate of false positive detections of 3.3%. This reduced accuracy is partly due to simplifying assumptions made by serial analysis. Furthermore, the detection of primary movements that spread over several notes is also a more difficult problem since the individual displacements due to key reaction force accumulate so that more secondary movement can be generated. The primary movements, however, do not grow larger in the same dimension.

To be usable for piano pedagogy, a sensor-based feedback system has to avoid unjustified corrective feedback. To do so, it can be necessary to adapt the sensitivity parameter according to the expectation of the system. If the student is required to perform a primary movement, a lower sensitivity parameter should be used, while a higher sensitivity parameter should be used when the student is instructed to avoid a primary movement.

Pedagogical applications: Starting with Deppe and his followers in the late 19th century, the importance of arm movements for piano technique was recognized [54, pp. 229–270]. Today many pedagogues acknowledge that both arm and finger movements are important [54, pp. 447–491]. Two pedagogical applications were developed that help the student to gain a better understanding and awareness of his arm movements: The first application provides auditive feedback on the student's movement. It is based on a movement notation that indicates movements with symbols in a traditional music score. In order to make a connection to existing piano pedagogical practices, existing movement notation were surveyed. We chose to support a movement notation system developed by the piano pedagogue S. Bernstein [9–11] since there exist exercise books that employ this notation [9, 10]. These exercises can now be supported with sensor-based feedback. Bernstein's movement notation was realized using our movement analysis methods. A user study with piano students of a music university showed that potential users think that the system is useful for learning technique. Furthermore, the movement detection accuracy is rated very high by the users, which underlines the quality achieved by our analysis methods.

The second pedagogical application visualizes the sensor data and allows the users to analyze and interpret the sensor signal. It visualizes sensor data, MIDI data, and the musical score and replays the audio/video recording of the performance. To simplify manual analysis, the performance of the same piece played by different persons can be synchronized so that video and sensor data can be compared. This makes it easy to spot differences where a closer examination may be beneficial. To help the users to cope with secondary movement in the sensor graphs the expected amounts of secondary movement are marked in the visualization.

Hand tracking: As discrete analysis depends on key reaction force, which is determined based on loudness information provided by a MIDI interface, it is necessary to know which hand has played the note. For this purpose two methods were introduced: one method uses MIDI data exclusively. The other method uses inertial sensor data in addition to MIDI data. Both methods are based on a series of two steps. In the first step, a received note-on event is assigned to the left or right hand. In the second step the note-on event is used to modify the estimated position of the hand. The methods scan the stream of MIDI data for simultaneously pressed keys that are located too far from each other to be played by one hand alone. Assuming that the hands do not cross over, these "unique notes" are assigned to the left or right hand. If a note is not a unique note, it is assigned to a hand based on the distance of the note to the current estimate of the hand position. Hand position is estimated with one Kalman filter for each hand. For this purpose, the pitch of a received note-on event is interpreted as an approximate measurement for the hand position along the keyboard. The method based on inertial sensors uses the movement measurements to continuously update the hand position even when no note-on event is reported. A computer-vision-based methods that was originally developed by Lefebvre-Albaret & Dalle [116] for Sign Language Recognition is evaluated for tracking hands during piano performance.

All three methods are effective to determine which hand has played a note. The best results are achieved with computer-vision-based method, which is closely followed by the sensor-based and the MIDI-based method. The main advantage of the sensor-based method is that it does not rely on a controlled environment (lighting, floor color, etc.), which is necessary when tracking hands visually. The main advantage of the MIDI-based method is that it does not depend on any additional hardware, making it applicable for assigning notes to staffs in context of music score notation software.

9.2. Possible improvements and new directions

More extensive measurement: In order to increase the accuracy of the pianist arm movement analysis, it would be beneficial to perform more extensive measurements of the factors that influence secondary movement. A potential candidate would be to determine which finger played a note. This has a distinct effect on the secondary movement in the arm as the point of application of the key-reaction force depends on the used finger.

Continuous key position measurement, which can be provided by the player pianos Yamaha Disklavier [184] and Bechstein CEUS [14], could be used to increase the accuracy. Key position measurement would allow determining when the finger makes contact with the key and when the key is fully depressed. This allows determining more accurately when the touch movement begins and is completed than the currently used estimate.

Lukowicz et al. proposed a method to measure muscle activity using force sensors to record changes in muscle shape [118]. This could be used to improve the estimation of secondary movement since the amount of secondary movement is reduced when the muscles that stabilize a joint are contracted.

Finger movement analysis: Capturing finger movements in piano playing is a delicate task. The current solutions to track fingers are not without problems: Data gloves interfere with playing movements; motion capture systems are prohibitively expensive and have the problem of marker occlusion. Therefore, this work refrained from measuring and analyzing finger movements.

Unobtrusive assessment of muscle tension: Approaches exist for measuring EMG signals as a basis for feedback on muscle tension for piano pedagogy [132, 155, 156]. However, EMG measurements are obtrusive since electrodes have to be placed on the player's skin. As muscle tension has an influence on the secondary movement that is generated by key-reaction forces (higher tension leads to less secondary movement), it could be possible to estimate the amount of muscle tension by examining the secondary movements based on an exact measurement of reaction force.

Other application areas: Beyond piano movement analysis we see further application areas for our methods: First, our methods could be used to provide sensor-based feedback for other tasks that have a significant amount of secondary movement. Examples are tasks that have large secondary movements due to large reaction forces (e. g., shooting a ball), tasks that are executed with relatively small primary movements (e. g., various forms of handcraft and instrument performance), and tasks where one wants to check if a part of the body is be kept still (e. g., gymnastic exercises). Second, our methods could be used to improve current gesture recognition techniques by removing secondary movement from the sensor signal so that secondary movement is not misinterpreted as an execution of a gesture or activity that was not actually performed. Consider, e. g., a sensor-equipped headset similar to the Talking Assistant described by Aitenbichler [1] with the following hypothetical gestural interface: The system makes a request via the headphones and asks for a yes or no decision. The user nods to signalize his approval to the system. When walking, the reaction forces that are transmitted via the foot to the body lead to secondary movement in the head. These reaction forces could be measured in the shoe or estimated using the accelerometer signals of sensors contained in the headset. By using our methods secondary head movements could be ignored so that false positive detections of the gesture can be avoided without requiring the user to exaggerate the gesture.

Movement analysis based on Labanotation: Labanotation (see Figure 9.1) is a method to notate human movement. It has been developed for notating dance and is used mainly for this purpose [69]. While many dance notations cover a particular dance style [68], Labanotation is claimed to be usable to notate human movement in general. In fact, Labanotation has already been used to notate movements in athletics and physiotherapy [69, p. 5]. The ability to determine whether a movement has been executed as indicated by a Labanotation score could enable applications in fields such as rehabilitation, gymnastics, and ergonomics. This can be seen as an extension of our approach of using Bernstein's movement notation to provide sensor-based feedback for piano pedagogy.

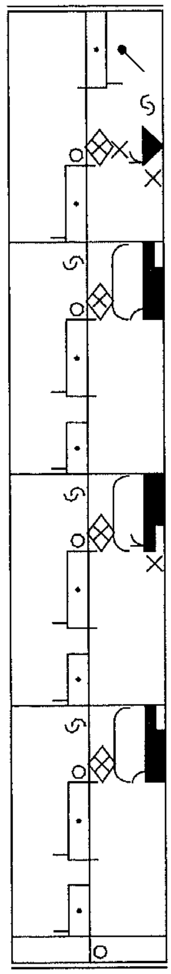


Figure 9.1.: Labanotation score of a Spanish dance [69, p. 195]. The score is read from bottom to top. The centerline represents the center of the body. Movements on the right side of the body are indicated right from the centerline [69]. An introduction to Labanotation can be found in [69].

Appendix A.

Estimation of secondary movement

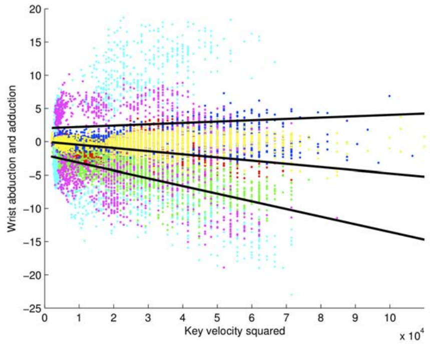
Contents

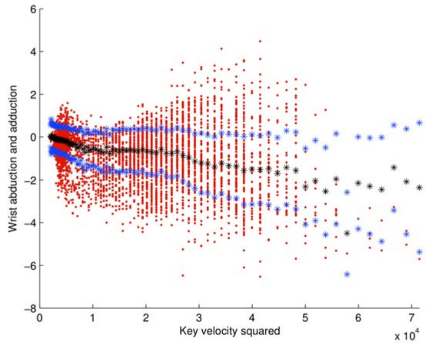
In the following, the estimations of secondary movement are shown for each of the main degrees of freedom of the human arm. The estimations were determined by discrete analysis for touch analysis as described in Sections 6.2 and 6.3. The following plots indicate measurements of secondary movement with colored dots. The color of a dot represents the touch type that was performed:

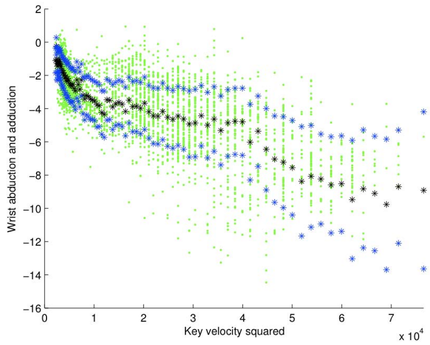
- Finger touches: red
- Hand touches: green
- Forearm touches: blue
- Pronation touches: cyan
- Supination touches: magenta
- Shoulder touches: yellow

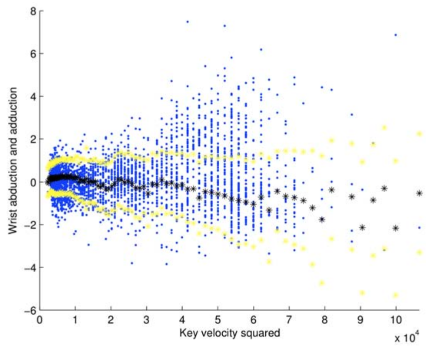
In the plots generated by the minimal model, where the estimated secondary movement depends only on the loudness information from the MIDI-enabled piano, three black lines indicate the mean (center line) and the mean ± the standard deviation (outer lines). In the plots generated by the full model, where the estimated secondary movement depends not only on the loudness information from the MIDI-enabled piano but also from the movements in other parts of the arm, black dots represent the estimated mean of secondary movement. Blue or yellow dots indicate the estimated mean ± the standard deviation.

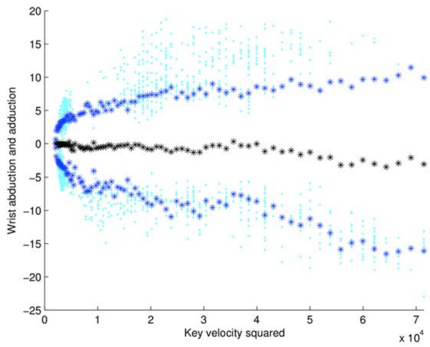
A.1. Wrist abduction and adduction

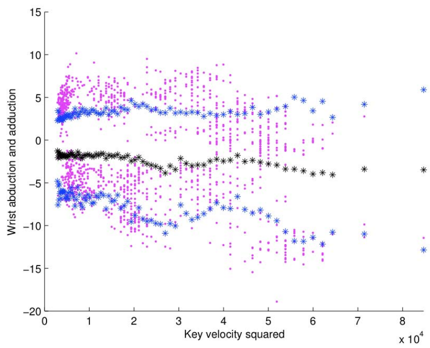


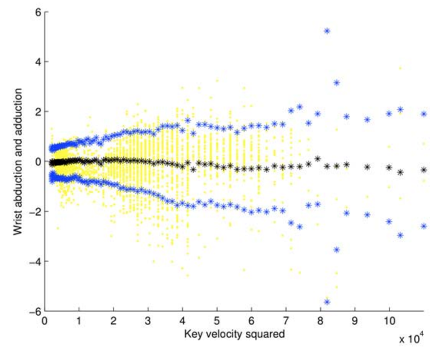




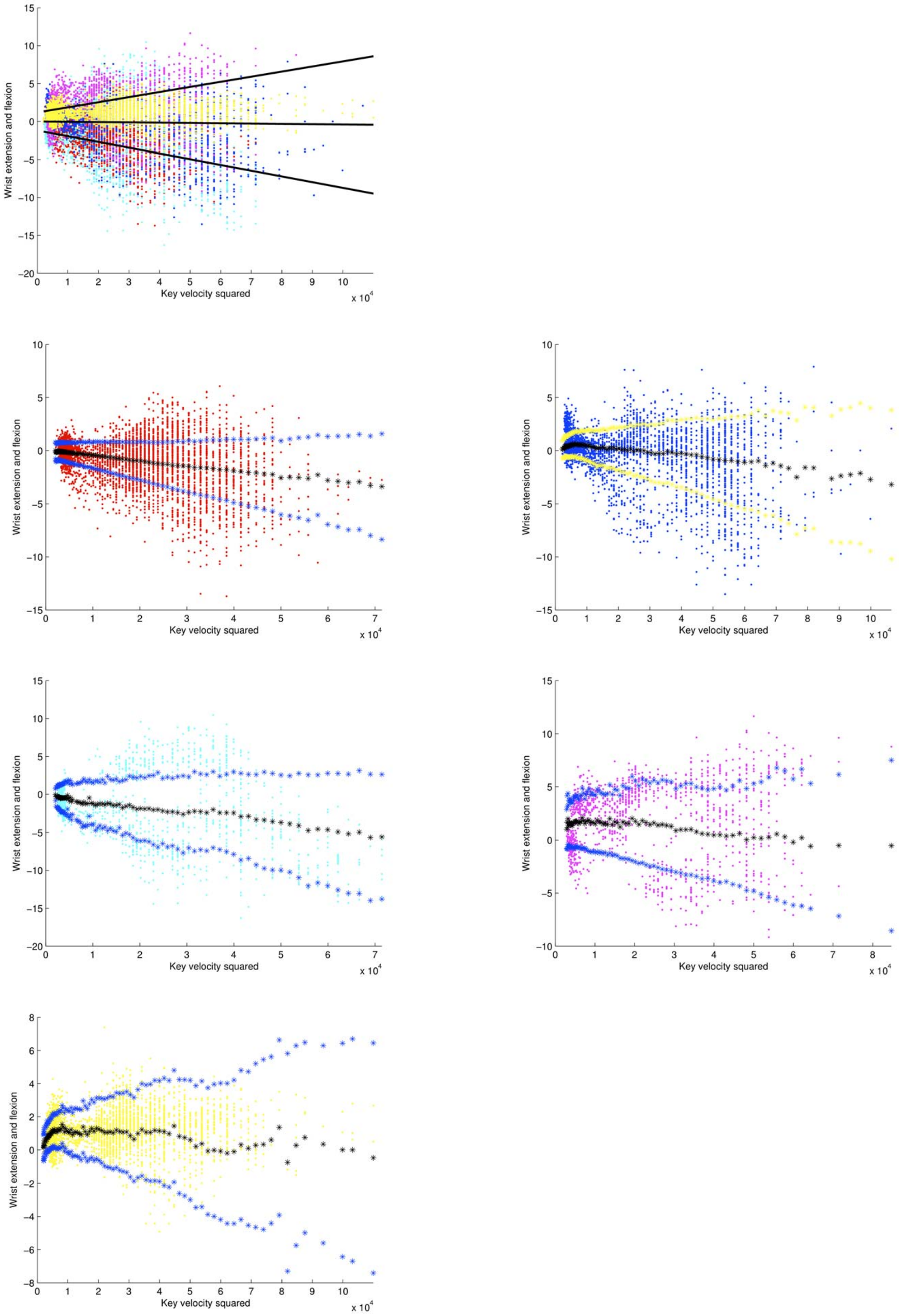




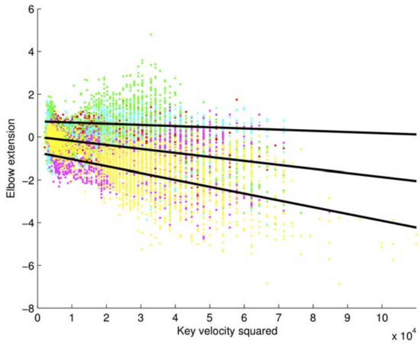


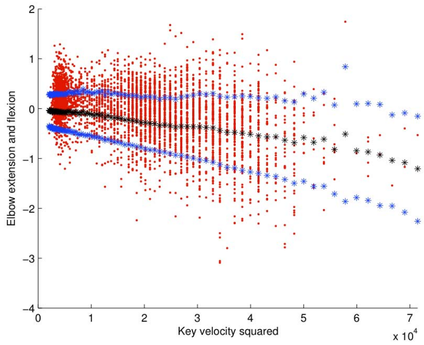


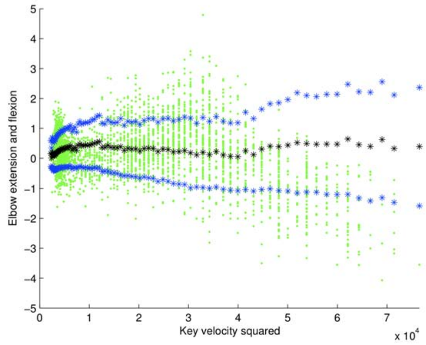
A.2. Wrist extension and flexion

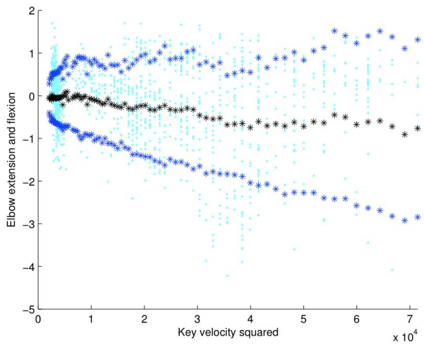


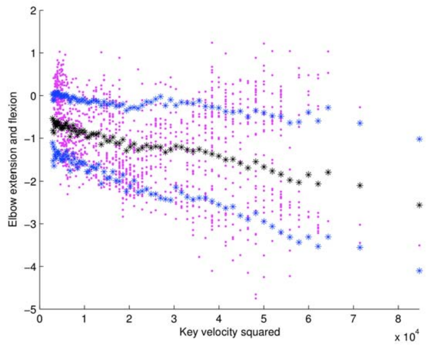
A.3. Elbow extension and flexion

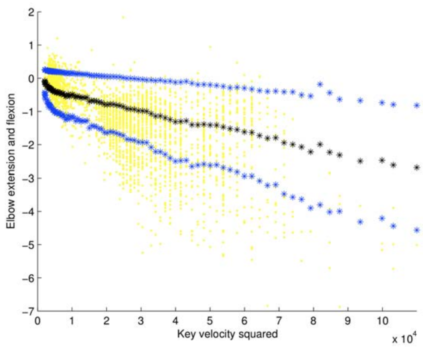




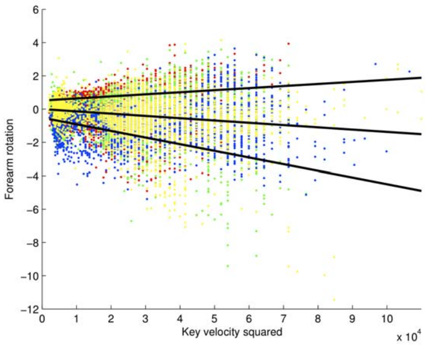


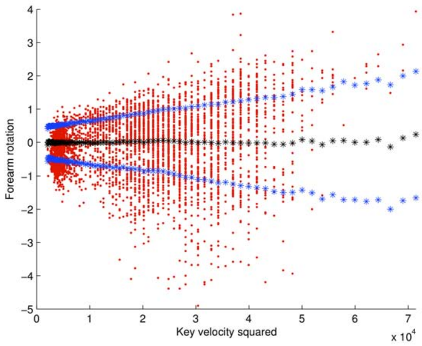


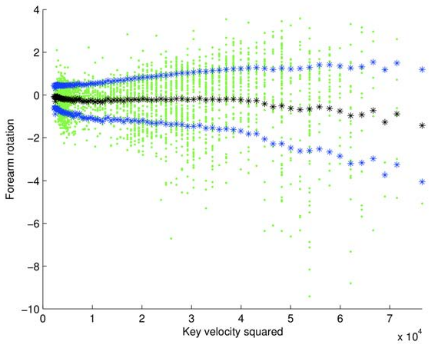


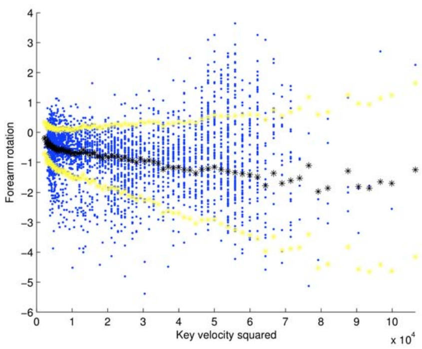


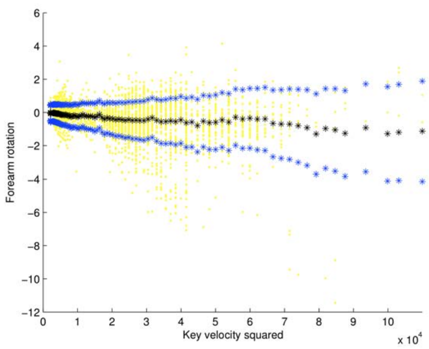
A.4. Forearm rotation



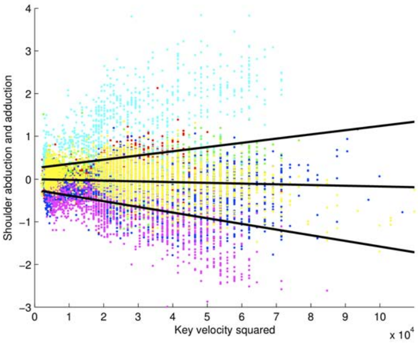


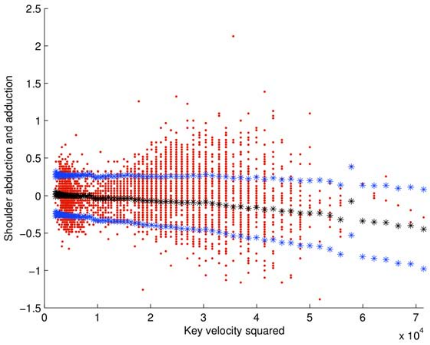


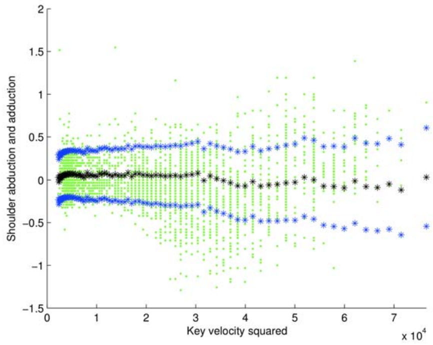


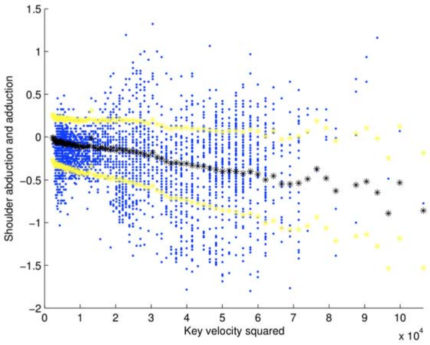


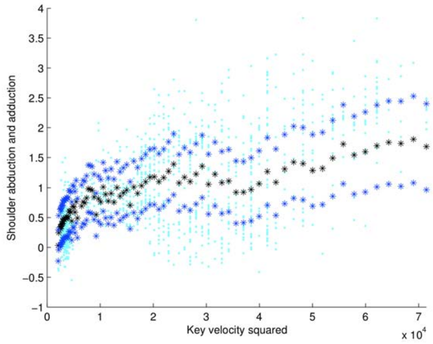
A.5. Shoulder abduction and adduction

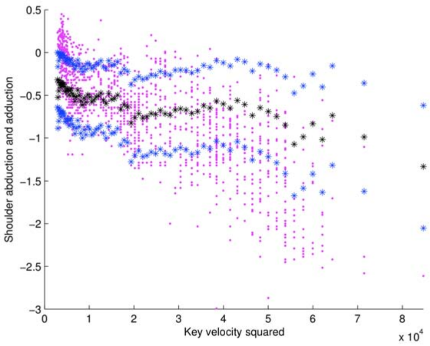


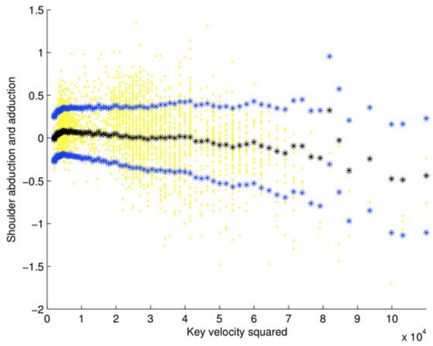




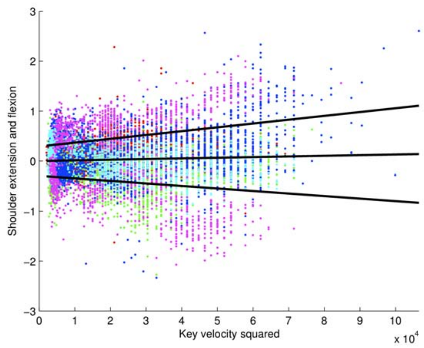


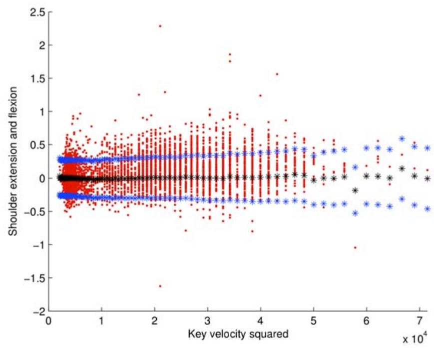


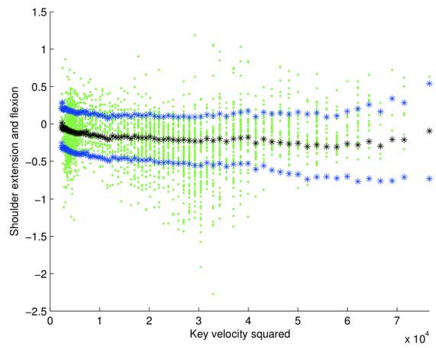


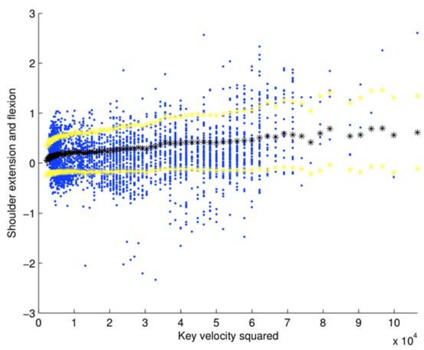


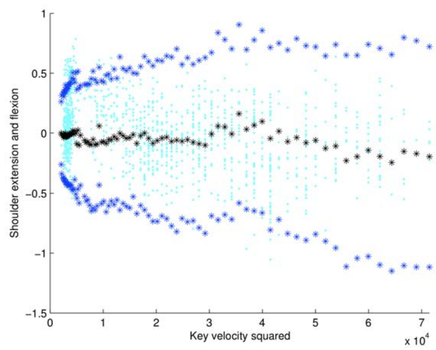
A.6. Shoulder extension and flexion

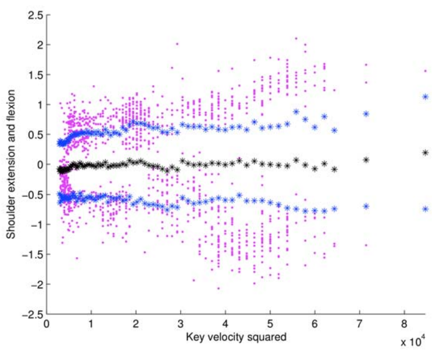




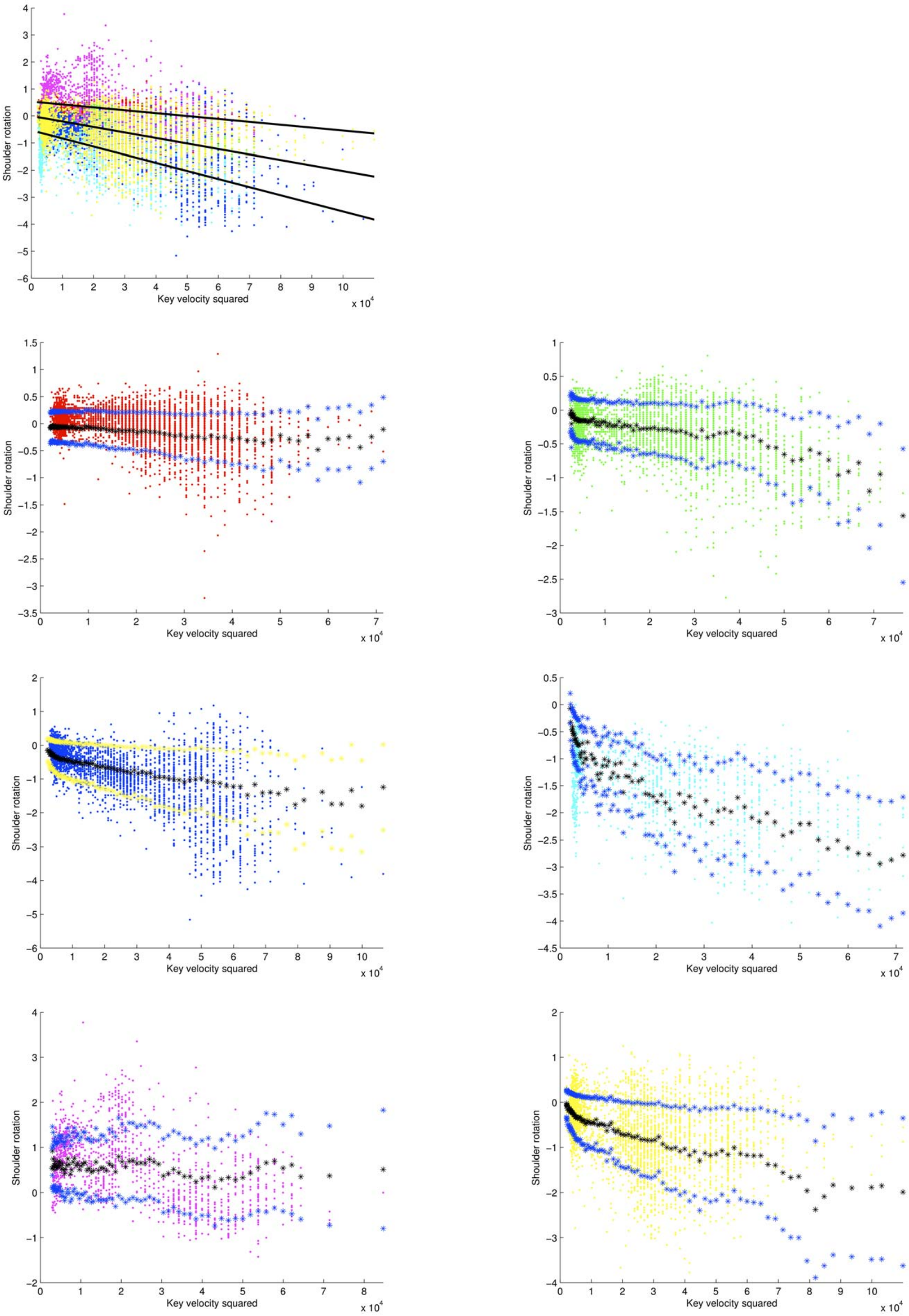








A.7. Shoulder rotation



Bibliography

[1] E. Aitenbichler. *System Support for Ubiquitous Computing.* PhD thesis, TU Darmstadt, 2006.

[2] E. Aitenbichler, F. Lyardet, A. Hadjakos, and M. Mühlhäuser. Fine-Grained Evaluation of Local Positioning Systems for Specific Target Applications. In *Ubiquitous Intelligence and Computing*, volume 5585 of *LNCS*, pages 236–250. Springer, 2009.

[3] S. Akinaga, M. Miura, N. Emura, and M. Yanagida. An algorithm to evaluate the appropriateness for playing scales on the piano. In *Proc. of the International Congress on Acoustics*, 2007.

[4] A. Askenfeld. Measurement of bow motion and bow force in violin playing. *Journal of the Acoustical Society of America*, 80, 1986.

[5] A. Baca. *Computers in Sports*, chapter Feedback systems, pages 43–67. WIT Press, 2008.

[6] E. J. Bach. *Die vollendete Klaviertechnik.* Breitkopf & Härtel, Leipzig, 2nd edition, 1960.

[7] F. J. Bejjani, L. Ferrara, N. Xu, C. M. Tomaino, L. Pavlidis, J. Wu, and J. Dommerholt. Comparison of three piano techniques as an implementation of a proposed experimental design. *Medical Problems of Performing Artists*, 4(3), 1989.

[8] S. Bériault, M. Côté, and P. Payeur. Volumetric modeling with multiple cameras for markerless motion capture in complex scenes. In *IEEE International Instrumentation and Measurement Technology Conference*, 2008.

[9] S. Bernstein. *Musi-Physi-Cality: Making a Physical Connection to Musical Feeling for the Beginning Pianist.* Seymour Bernstein Music, 1991.

[10] S. Bernstein. *Twenty Lessons In Keyboard Choreography.* Seymour Bernstein Music, 1991.

[11] S. Bernstein. *With your own two hands: self-discovery through music.* G. Schirmer, 1991.

[12] C. M. Bishop. *Pattern recognition and machine learning.* Springer, 2006.

[13] J. Borchers, A. Hadjakos, and M. Mühlhäuser. MICON: a music stand for interactive conducting. In *New Interfaces for Musical Expression 2006*, 2006.

[14] Bösendorfer. CEUS technical characteristics. `http://www.boesendorfer.com/en/technical-characteristics.html`, June 2011.

[15] A. Bouënard, M. M. Wanderley, and S. Gibet. Analysis of timpani preparatory gesture parameterization. In *Gesture Workshop 2009*, 2009.

[16] D. A. Bowman, E. Kruijff, J. J. LaViola, Jr., and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison Wesley, 2004.

[17] M. Brandão, G. Wiggins, and H. Pain. Computers in music education. In *Proceedings of the AISB'99 Symposium on Musical Creativity*, pages 82–88, 1999.

[18] R. M. Breithaupt. *Die Natürliche Klaviertechnik. Band II. Die Grundlagen des Gewichtspiels*. C. F. Kahnt Nachfolger, Leipzig, 1909.

[19] R. M. Breithaupt. *Die Natürliche Klaviertechnik. Band I. Handbuch der modernen Methodik und Spielpraxis*. C. F. Kahnt Nachfolger, 3rd edition, 1911.

[20] R. M. Breithaupt. *Die Natürliche Klaviertechnik. Band III. Praktische Studien. Heft I*. C. F. Kahnt Nachfolger, 1917.

[21] A. Burns and M. Wanderley. Visual methods for the retrieval of guitarist fingering. In *New Interfaces for Musical Expression 2006*, 2006.

[22] O. Cakmakci, F. Bèrard, and J. Coutaz. An augmented reality based learning assistant for electric bass guitar. In *Proc. of the 10th International Conference on Human-Computer Interaction, Crete, Greece*, 2003.

[23] B. Calais-Germain. *Anatomy of Movement*. Eastland Press, 2007.

[24] A. P. Carrillo and J. Bonada. The bowed tube: a virtual violin. In *New Interfaces for Musical Expression 2010*, 2010.

[25] R. Carroll and D. Ruppert. *Transformation and weighting in regression*. Chapman & Hall/CRC, 1988.

[26] V. Chandola, A. Banerjee, and V. Kumar. Outlier detection: A survey. *ACM Computing Surveys*, pages 1–72, 2009.

[27] E. Chew and X. Wu. Separating voices in polyphonic music: A contig mapping approach. In *Computer Music Modeling and Retrieval: Second International Symposium*, pages 1–20. Springer, 2004.

[28] I.-S. Chung, J. Ryu, N. Ohnishi, B. Rowen, and J. Headrich. Wrist motion analysis in pianists. *Medical Problems of Performing Artists*, 7(1), 1992.

[29] A. Cont. Antescofo: Anticipatory synchronization and control of interactive parameters in computer music. In *International Computer Music Conference 2008*, 2008.

[30] J. Craig. *Introduction to Robotics: Mechanics and Control*. Pearson, 3rd edition, 2005.

[31] A. Cuvelier, M. Decoust, and B. Maton. Etude systématique des caractéristiques physiologiques et phsicologiques des musiciéns. Technical report, Etablissements Georges Pompidou, Département pédagogique, Département instruments et voix, 1977.

[32] R. Dannenberg. An on-line algorithm for real-time accompaniment. In *International Computer Music Conference 1984*, 1984.

[33] R. B. Dannenberg, M. Sanchez, A. Joseph, R. Joseph, R. Saul, and P. Capell. Results from the piano tutor project. In *Proceedings of the Fourth Biennial Arts and Technology Symposium*, pages 143–150, 1993.

[34] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(8):800–810, 2001.

[35] A. Devices. $\pm 300°/s$ *Single Chip Yaw Rate Gyro with Signal Conditioning ADXRS300*, 2004.

[36] A. Devices. *Small, Low Power, 3-Axis* $\pm 3\,g$ *iMEMS Accelerometer ADXL330*, 2006.

[37] S. Dixon, W. Goebl, and G. Widmer. The performance worm: Real time visualization of expression based on langner's tempo-loudness animation. In *International Computer Music Conference 2002*, pages 361–364, 2002.

[38] K. Essl. *Zwischen Experiment und Kommerz. Zur Ästhetik elektronischer Musik*, chapter Wandlungen der elektroakustischen Musik, pages 37–84. Mille Tre, 2007.

[39] S. Ferguson. Learning musical instrument skills through interactive sonification. In *New Interfaces for Musical Expression 2006*, 2006.

[40] S. Ferguson, A. Moere, and D. Cabrera. Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians. In *IEEE Proceedings of the International Conference on Information Visualisation*, pages 97–102, 2005.

[41] V. F. Ferrario, C. Macri, E. Biffi, P. Pollice, and C. Sforza. Three-dimensional analysis of hand and finger movements during piano playing. *Medical Problems of Performing Artists*, 22(1), 2007.

[42] D. Fober, S. Letz, and Y. Orlarey. Vemus – feedback and groupware technologies for music instrument learning. In *Sound and Music Computing 2007*, 2007.

[43] D. Fober, S. Letz, Y. Orlarey, A. Askenfeld, K. Falkenberg, and E. Schoonderwaldt. IMUTUS: An interactive music tuition system. In *Sound and Music Computing 2004*, pages 97–103, 2004.

[44] D. Focht. *Ergonomics for Therapists*, chapter Lifting Analysis. Elsevier, 2008.

[45] Y. C. Fung. *Biomechanics: motion, flow, stress, and growth.* Springer, 1990.

[46] Y. C. Fung. *Biomechanics: mechanical properties of living tissues.* Springer, 2nd edition, 1993.

[47] Y. C. Fung. *Biomechanics: circulation.* Springer, 2nd edition, 1997.

[48] S. Furuya and H. Kinoshita. Roles of proximal-to-distal sequential organization of the upper limb segments in striking the keys by expert pianists. *Neuroscience Letters*, pages 264–269, 2007.

[49] S. Furuya and H. Kinoshita. Expertise-dependent modulation of muscular and non-muscular torques in multi-joint arm movements during piano keystroke. *Neurosience*, 156:390–402, 2008.

[50] S. Furuya and H. Kinoshita. Organization of the upper limb movement for piano key-depression differs between expert pianists and novice players. *Exp Brain Res*, 185:581–593, 2008.

[51] S. Furuya, R. Osu, and H. Kinoshita. Effective utilization of gravity during arm downswing keystrokes by expert pianists. *Neurosience*, 164:822–831, 2009.

[52] J. Gat. *The Technique of Piano Playing.* Collet's Holding, London, 2nd edition, 1965.

[53] S. Geman and C.-R. Hwang. Nonparametric maximum likelihood estimation by the method of sieves. *The Annals of Statistics*, 10(2), 1982.

[54] R. R. Gerig. *Famous Pianists and Their Technique.* Indiana University Press, 2007.

[55] O. Gillet and G. Richard. Automatic transcription of drum sequences using audiovisual features. In *ICASSP'05*, 2005.

[56] A. Gkiokas, K. Perifanos, and S. Nikolaidis. Real-time detection and visualization of carinet bad sounds. In *DAFx-08*, 2008.

[57] W. Goebl, R. Bresin, and A. Galembo. Touch and temporal behavior of grand piano actions. *Journal of the Acoustical Society of America*, 118(2):1154–1165, 2005.

[58] W. Goebl and C. Palmer. Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research*, 186(3):471–479, 2008.

[59] W. Goebl and G. Widmer. Unobstrusive practice tools for pianists. In *Proceedings of the 9th International Conference on Music Perception and Cognition*, 2006.

[60] D. O. Gorodnichy and A. Yogeswaran. Detection and tracking of pianist hands and fingers. In *Proc. of the The 3rd Canadian Conference on Computer and Robot Vision*, 2006.

[61] M. Gövercin, I. M. Missala, M. Marschollek, and E. Steinhagen-Thiessen. Virtual rehabilitation and telerehabilitation for the upper limb: A geriatric review. *GeroPsyhc*, 23(2):79–90, 2010.

[62] T. Großhauser. Low force pressure measurement: Pressure sensor matrices for gesture analysis, stiffness recognition and augmented instruments. In *New Interfaces for Musical Expression 2008*, 2008.

[63] T. Großhauser, U. Großekathöfer, and T. Hermann. New sensors and pattern recognition techniques for string instruments. In *New Interfaces for Musical Expression 2010*, 2010.

[64] T. Großhauser and T. Hermann. Augmented haptics—an interactive feedback system for musicians. In *Proceedings of the 4th International Conference on Haptic and Audio Interaction Design*, pages 100–108. Springer, 2009.

[65] T. Großhauser and T. Hermann. The sonified music stand – an interactive sonification system for musicians. In *Sound and Music Computing 2009*, 2009.

[66] T. Großhauser and R. Trautner. Everyday sensor system for music instruments: possibilities and usage in daily musical instrument playing, exercising and teaching. In *Proceedings of the 5th conference on Applied electromagnetics, wireless and optical communications*, pages 123–126, 2007.

[67] E. Guaus, T. Ozaslan, E. Palacios, and J. L. Arcos. A left hand gesture caption system for guitar based on capacitive sensors. In *New Interfaces for Musical Expression 2010*, 2010.

[68] A. H. Guest. *Choreo-Graphics: A Comparison of Dance Notation Systems from the Fifteenth Century to the Present.* Gordon and Breach, 1998.

[69] A. H. Guest. *Labanotation: The System of Analyzing and Recording Movement.* Routledge, 4th edition, 2005.

[70] N. Habili, C. C. Lim, and A. Moini. Segmentation of the face and hands in sign language video sequences using color and motion cues. *IEEE Trans. Circuits Syst. Video Techn.*, 14(8), 2004.

[71] A. Hadjakos. MICON: a music stand for interactive conducting. Master's thesis, TU Darmstadt, 2006.

[72] A. Hadjakos. Feinevaluierung des Probabilistischen Arm Modells zur Analyse von Klavierspielbewegungen. In *Workshop für Innovative Computerbasierte Musikinterfaces 2009*, 2009.

[73] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. The elbow piano: Sonification of piano playing movements. In *New Interfaces for Musical Expression 2008*, pages 285–288, 2008.

[74] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. Potential use of inertial measurement sensors for piano teaching systems: Motion analysis of piano playing patterns. In *Proceedings of the 4th i-Maestro Workshop on Technology-Enhanced Music Education*, 2008.

[75] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. Syssomo: A pedagogical tool for analyzing movement variants between different pianistst. In E. Ruffaldi and M. Fontana, editors, *Enactive08 Proceedings*, pages 75–80, Pisa, 2008. Edizioni ETS.

[76] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. Zu einer automatischen Erkennung von Spielbewegungen am Klavier. In U. Lucke, M. C. Kindsmüller, S. Fischer, M. Herczeg, and S. Seehusen, editors, *Workshop Proceedings der Tagungen Mensch & Computer 2008, DeLFI 2008 und Cognitive Design 2008*, pages 97–99, Berlin, 2008. Logos Verlag.

[77] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. Aiding the development of piano gesture recognition with a gesture enriched instrument. In *Gesture Workshop 2009*, 2009.

[78] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. Probabilistic model of pianists' arm touch movements. In *New Interfaces for Musical Expression 2009*, pages 7–12, 2009.

[79] A. Hadjakos and F. Lefebvre-Albaret. Three methods for pianist hand assignment. In *Sound and Music Computing 2009*, 2009.

[80] A. Hadjakos and M. Mühlhäuser. Analysis of piano playing movements spanning multiple touches. In *New Interfaces for Musical Expression 2010*, 2010.

[81] D. C. Harding, K. D. Brandt, and B. M. Hilberry. Minimization of finger joint forces and tenson tensions in pianists. *Medical Problems of Performing Artists*, 4(3), 1989.

[82] D. C. Harding, K. D. Brandt, and B. M. Hillberry. Finger joint force minimization in pianists using optimization techniques. *Biomechanics*, 26(12):1403–1412, 1993.

[83] H. Hatze. The meaning of the term 'biomechanics'. *J. Biomechanics*, 7:189–190, 1974.

[84] W. Herzog. The biomechanics of muscle contraction: or firing biomechanics research. In R. Magjarevic, J. H. Nagel, N. A. Abu Osman, F. Ibrahim, W. A. B. Wan Abas, H. S. Abdul Rahman, and H.-N. Ting, editors, *4th Kuala Lumpur International Conference on Biomedical Engineering 2008*, volume 21 of *IFMBE Proceedings*. Springer, 2008.

[85] W. Herzog and P. Binding. *Mathematically Indeterminate Systems*, chapter 4.8, pages 533–545. John Wiley & Sons, 1999.

[86] I. Hmelnitsky and N. Nettheim. Weight-bearing manipulation: A neglected area of medical science relevant to piano playing and overuse syndrome. *Medical Hypotheses*, 23:209–217, 1987.

[87] R. R. Hocking. *Methods and Applications of Linear Models.* Wiley, 2nd edition, 2003.

[88] P. Hodgson. *Motion study and violin bowing.* L.H. Lavender & Co, London, 1934.

[89] H. H. Hoos, K. A. Hamel, K. Renz, and J. Kilian. The guido music notation format—a novel approach for adequately representing score-level music. In *International Computer Music Conference 1998*, pages 451–454, 1998.

[90] D. M. Howard, G. F. Welch, J. Brereton, E. Himonides, M. DeCosta, J. Williams, and A. W. Howard. Winsingad: a real-time display for the singing studio. *Logoped Phoniatr Vocol*, 29(135–144), 2004.

[91] K. Huang, E. Y.-L. Do, and T. Starner. Pianotouch: A wearable haptic piano instruction system for passive learning of piano skills. In *ISWC 2008*, 2008.

[92] D. T. G. Huyn. *Human Activity Recognition with Wearable Sensors.* PhD thesis, TU Darmstadt, 2008.

[93] Immersion. *CyberGlove Reference Manual*, 1998.

[94] Invensense. *Integrated Dual-Axis Gyro IDG-300*, 2006.

[95] N. Iwami and M. Miura. A Support System for Basic Practice of Playing the Drums. In *International Computer Music Conference 2007*, pages 364–367, 2007.

[96] J. D. Jobson. *Applied Multivariate Data Analysis: Regression and Experimental Design.* Springer, 1999.

[97] K. Johnen. *Wege zur Energetik des Klavierspiels.* Mitteldeutscher Verlag Halle, 1951.

[98] A. Jordanous. Voice separation in polyphonic music: A data-driven approach. In *International Computer Music Conference 2008*, 2008.

[99] R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, pages 35–45, 1960.

[100] G. Kamen. *Reserach Methods in Biomechanics*, chapter Electromyographic Kinesiology. Human Kinetics, 2004.

[101] B. Kemp, A. Janssen, and B. van der Kamp. Body position can be monitored in 3D using miniature accelerometers and earth-magnetic field sensors. *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, 109(6):484–488, 1998.

[102] C. Kerdvibulvech and H. Saito. Guitarist fingertip tracking by integrating a bayesian classifier into particle filters. *Advances in Human-Computer Interaction*, 2008.

[103] C. Kerdvibulvech and H. Saito. Markerless guitarist fingertip detection using a bayesian classifier and a template matching for supporting guitarists. In *10th IEEE Virtual Reality International Conference (VRIC 08)*, 2008.

[104] J. Kilian and H. Hoos. Voice separation—a local optimisation approach. In *Proc. International Conference on Music Information Retrieval–ISMIR'2002*, 2002.

[105] J. Kim, F. Cordier, and N. Magnenat-Thalmann. Neural network-based violinist's hand animation. In *Computer Graphics International*, pages 37–41, 2000.

[106] H. Kinoshita, S. Furuya, T. Aoki, H. Nakahara, and E. Altenmuller. Characteristics of keystroke force in the piano. In *International Society of Biomechanics XXI congress*, 2007.

[107] A. Klapuri and M. E. Davy. *Signal Processing Methods for Music Transcription.* Springer, 2006.

[108] G. Kochevitsky. *The Art of Piano Playing: a scientific approach.* Summy-Birchard Music, 1967.

[109] R. Koenker. *Quantile Regression.* Econometric Society Monographs. Cambridge University Press, 2005.

[110] R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[111] Korg. *Korg X5 Music Synthesizer AI2 Synthesis System Bedienungshandbuch*, 1994.

[112] D. Koutsonanos, M. Papadogiorgaki, E. Kalapanidas, D. Tzovaras, C. Davarakis, and M. Strintzis. Enhancing Electronic Music Tuition Systems with Virtual Reality Tools. In *5th Open Workshop of MUSICNETWORK: Integration of Music in Multimedia Applications, Vienna, Austria*, 2005.

[113] N. Kugimoto, R. Miyazono, K. Omori, T. Fujimura, S. Furuya, H. Katayose, H. Miwa, and N. Nagata. CG animation for piano performance. In *SIGGRAPH '09: Posters*, New York, NY, USA, 2009. ACM.

[114] C.-L. Lai, R.-C. Chan, T.-F. Yang, and I.-W. Penn. EMG changes during graded isometric exercise in pianists: Comparison with non-musicians. *J Chin Med Assoc*, 71(11), 2008.

[115] Q. Le, A. Smola, and S. Canu. Heteroscedastic Gaussian process regression. In *Proceedings of the 22nd international conference on Machine learning*, 2005.

[116] F. Lefebvre-Albaret and P. Dalle. Body posture estimation in sign language video. In *Gesture Workshop 2009*, 2009.

[117] W. LeVine and J. Irvine. In vivo emg biofeedback in violin and viola pedagogy. *Applied Psychophysiology and Biofeedback*, 9(2):161–168, 1984.

[118] P. Lukowicz, F. Hanser, C. Szubski, and W. Schobersberger. Detecting and interpreting muscle activity with wearable force sensors. In K. Fishkin, B. Schiele, P. Nixon, and A. Quigley, editors, *Pervasive Computing*, Lecture Notes in Computer Science, pages 101–116. Springer, 2006.

[119] S. Madsen and G. Widmer. Separating voices in midi. In *Proceedings of the 9th International Conference in Music Perception and Cognition (ICMPC2006)*, pages 22–26, 2006.

[120] E. Maestre, J. Janer, M. Blaauw, A. Pérez, and E. Guaus. Acquisition of violin instrumental gestures using a commercial emf tracking device. In *Proceedings of the 2007 International Computer Music Conference, Copenhagen, Denmark*, 2007.

[121] F. Mahmoudi and M. Parviz. Visual hand tracking algorithms. In *Geometric Modeling and Imaging–New Trends*, 2006.

[122] C. Marek. *Lehre des Klavierspiels*. Atlantis, Zurich, 2nd edition, 1977.

[123] T. Matthay. *The Act of Touch in All Its Diversity: an Analysis and Synthesis of Pianoforte Tone-Production*. Bosworth, London, 1903.

[124] P. McLeod and G. Wyvill. Visualization of musical pitch. In *Computer Graphics International 2003*, 2003.

[125] K. A. McMillen. Stage-worthy sensor bows for stringed instruments. In *New Interfaces for Musical Expression 2008*, 2008.

[126] A. McPherson and Y. Kim. Augmenting the acoustic piano with electromagnetic string actuation and continuous key position sensing. In *New Interfaces for Musical Expression 2010*, 2010.

[127] A. Menache. *Understanding motion capture for computer animation and video games*. Morgan Kaufmann Pub, 2000.

[128] A. Migeon and A.-E. Lenel. *Modern Sensors Handbook*, chapter Accelerometers and Inclinometers, pages 193–244. ISTE, London, 2007.

[129] A. Migeon and A.-E. Lenel. *Modern Sensors Handbook*, chapter Solid State Gyroscopes and Navigation, pages 395–432. ISTE, London, 2007.

[130] S. Mitra and T. Acharya. Gesture recognition: A survey. *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(3):311–324, 2007.

[131] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 2006.

[132] R. Montes, M. Bedmar, and M. S. Martin. Emg biofeedback of the abductor pollicis in piano performance brevis. *Biofeedback and Self-Regulation*, 18(2), 1993.

[133] J. Mora, W.-S. Lee, G. Comeau, S. Shirmohammadi, and A. E. Saddik. Assisted piano pedagogy through 3d visualization of piano playing. In *HAVE 2006 – IEEE International Workshop on Haptic Audio Visual Environments and their Application*, pages 157–160, 2006.

[134] S. Morita, N. Emura, M. Miura, S. Akinaga, and M. Yanagida. Feature exaggeration in scale performance on the piano. In *Acoustics08*, 2008.

[135] M. Mukai, N. Emura, M. Miura, and M. Yanagida. Generation of suitable phrases for basic training to overcome weak points in playing the piano. In *Proc. of International Congress on Acoustics*, 2007.

[136] N. Nagata, N. Kugimoto, R. Miyazono, K. Omori, T. Fujimura, S. Furuya, H. Katayose, and H. Miwa. CG animation for piano performance. In *Proc. 15th Japan-Korea Joint Workshop on Frontiers of Computer Vision (FCV2009)*, pages 302–307, 2009.

[137] K. Ng, T. Weyde, O. Larkin, K. Neubarth, T. Koerselman, and B. Ong. 3d augmented mirror: A multimodal interface for string instrument learning and teaching with gesture support. In *ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces*, 2007.

[138] S. Nicolls. Twenty-first century piano. In *New Interfaces for Musical Expression 2009*, 2009.

[139] B. M. Nigg and W. Herzog, editors. *Biomechanics of the musculo-skeletal system.* John Wiley & Sons, 1999.

[140] O. Ortmann. *The Physical Basis of Piano Touch and Tone: An Experimental Investigation of the Effect of the Player's Touch Upon the Tone of the Piano.* Kegan Paul, Trenc, Trubner & Co., Ltd., London; J. Curwen & Sons, London, Ltd; E. P. Dutton & Co., New York, 1925.

[141] O. Ortmann. *The Physiological Mechanics of Piano Technique.* K. Paul, Trench, Trubner & Co., 1929.

[142] C. Oshima, K. Nishimoto, and N. Hagita. A piano duo support system for parents to lead children to practice musical performances. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(2), 2007.

[143] C. Oshima, K. Nishimoto, and A. Konagaya. Toward computer supported piano lesson to opportunely advance to the creative stage. In *Proc. Artificial Intelligence and Soft Computing (ASC2001)*, pages 85–92, 2001.

[144] J. Paradiso and N. Gershenfeld. Musical applications of electric field sensing. *Computer Music Journal*, pages 69–89, 1997.

[145] D. Parlitz, T. Peschel, and E. Altenmüller. Assessment of dynamic finger forces in pianists: Effects of training and expertise. *Journal of Biomechanics*, 31:1063–1067, 1998.

[146] C. Peiper, D. Warden, and G. Garnett. An interface for real-time classification of articulations produced by violin bowing. In *New Interfaces for Musical Expression 2003*, pages 192–196, 2003.

[147] G. Percival, Y. Wang, and G. Tzanetakis. Effective use of multimedia for computer-assisted musical instrument tutoring. In *Proceedings of the international workshop on Educational multimedia and multimedia education*, 2007.

[148] C. Poepel and R. B. Dannenberg. Audio signal driven sound synthesis. In *International Computer Music Conference 2005*, 2005.

[149] H. Pohl and A. Hadjakos. Dance pattern recognition using dynamic time warping. In *Sound and Music Computing 2010*, 2010.

[150] C. Raphael. Current directions with musical plus one. In *Sound and Music Computing 2009*, 2009.

[151] N. Rasamimanana, F. Guedy, N. Schnell, J.-P. Lambert, and F. Bevilacqua. Three pedagogical scenarios using the sound and gesture lab. In *4th i-Maestro Workshop on Technology-Enhanced Music Education*, 2008.

[152] N. H. Rasamimanana, E. Fléty, and F. Bevilacqua. Gesture analysis of violin bow strokes. In *Gesture in Human-Computer Interaction*, Lecture Notes in Computer Science. Springer, 2006.

[153] K. Renz. An improved algorithm for spacing a line of music. In *International Computer Music Conference 2002*, 2002.

[154] H. Riehle, R. Fritsch, and H. Gärtner. A dynamometric anaylsis of piano playing. In *ISBS'98 Proceedings*, 1998.

[155] K. Riley. New views of piano playing through motion analysis and midi technology. *Journal of Technology in Music Learning*, 2007.

[156] K. Riley, E. E. Coons, and D. Marcarian. The use of multimodal feedback in retraining complex technical skills of piano performance. *Medical Problems of Performing Artists*, 20(2), 2005.

[157] S. Ripka and P. Ripka. *Modern Sensors Handbook*, chapter Magnetic Sensors, pages 433–475. ISTE, London, 2007.

[158] M. Robine, G. Percival, and M. Lagrange. Analysis of saxophone performance for computer-assisted tutoring. In *International Computer Music Conference 2007*, volume 2, pages 381–384, 2007.

[159] D. Roetenberg, H. J. Luinge, C. T. M. Baten, and P. H. Veltnik. Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation. *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 13(3):395–405, 2005.

[160] R. A. Schmidt and T. D. Lee. *Motor control and learning: a behavioral emphasis.* Human Kinetics, 4 edition, 2005.

[161] E. Schoonderwaldt, N. Rasamimanana, and F. Bevilacqua. Combining accelerometer and video camera: Reconstruction of bow velocity profiles. In *New Interfaces for Musical Expression 2006*, 2006.

[162] E. Schoonderwaldt and M. Wanderley. Visualization of bowing gestures for feedback: The hodgson plot. In *4th i-Maestro Workshop on Technology-Enhanced Music Education*, 2007.

[163] W. Sesink. *Verhältnis von Allgemeiner Didaktik zur Mediendidaktik*, chapter Mediendidaktisch leitende Interessen und Perspektiven. Hagen: Fernuniversität, 2005.

[164] C. Sforza, C. Macri, M. Turci, G. Grassi, and V. F. Ferrario. Neuromuscular patterns of finger movements during piano playing. definition of an experimental protocol. *Ital J Anat Embryol.*, 108(4):211–222, 2003.

[165] G. Smith and A. Johnston. Interactive software for guitar learning. In *Proceedings of the Autralasian Computer Music Conference*, 2007.

[166] S. W. Smoliar, J. A. Waterworth, and P. R. Kellock. pianoforte: A system for piano education beyond notation literacy. In *MULTIMEDIA '95: Proceedings of the Third ACM International Conference on Multimedia*, 1995.

[167] O. Spaniol. *Computers in Sports*, chapter Preface, pages xv–xvi. WIT Press, 2008.

[168] M. D. Springer. *The Algebra of Random Variables.* John Wiley & Sons, 1979.

[169] J. Stange-Elbe. *Elektroakustische Musik*, chapter Das andere Musikinstrument. Von elektrischen Spielinstrumenten zum Synthesizer. Laaber-Verlag, 2002.

[170] M. Stikic. *Towards Less Supervision for Scalable Recognition of Daily Activities.* PhD thesis, TU Darmstadt, 2009.

[171] D. Trueman and P. Cook. Bossa: The deconstructed violin reconstructed. *Journal of New Music Research*, 29(2):121–130, 2000.

[172] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *IEEE Trans. on Circuits and Systems for Video Technology*, 18(11), 2008.

[173] H.-J. Uchdorf. *Praktisches Handbuch für Klavierbauer.* VEB Fachbuchverlag, Leipzig, 1985.

[174] M. Varró. *Der Lebendige Klavierunterricht: Seine Methodik und Psychologie.* N. Simrock, Hamburg, London, 4th edition, 1958.

[175] B. Vercoe. The synthetic performer in the context of live performance. In *International Computer Music Conference 1984*, 1984.

[176] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *Proc. Graphicon-2003*, 2003.

[177] Y. Wang, B. Zhang, and O. Schleusing. Educational violin transcription by fusing multimedia streams. In *Emme '07: Proceedings of the international workshop on Educational multimedia and multimedia education*, pages 57–66, New York, NY, USA, 2007. ACM.

[178] G. Welch and G. Bishop. An introduction to the kalman filter. Technical report, University of North Carolina, 1995.

[179] B. Wetz. Tasteninstrumente. *Medizinische Probleme bei Musikern*, 1995.

[180] H. Wilmers. Bowsense – a minimalistic approach to wireless motion sensing. In *Linux Audio Conference 2009*, 2009.

[181] F. G. Wolf, M. S. Keane, K. D. Brandt, and B. M. Hilberry. An investigation of finger joint and tendon forces in experienced pianists. *Medical Problems of Performing Artists*, 8(3), 1993.

[182] K. Wolters. *Das Klavier.* Hallawag Verlag, Bern, 1975.

[183] B. Wristen, S. Evans, and N. Stergiou. Sight-reading versus repertoire performance on the piano: A case study using high-speed motion analysis. *Medical Problems of Performing Artists*, 21(1), 2006.

[184] Yamaha. *Disklavier PRO / Mark IV Series Advanced Operating Manual*, 2004.

[185] X. Yan and X. Su. *Linear regression analysis: theory and computing.* World Scientific, 2009.

[186] M.-J. Yoo and I.-K. Lee. Interactive music summarization based on interval distance profile. In *ISMIR 2007*, 2007.

[187] M.-J. Yoo and I.-K. Lee. Pattern-specific piano exercise retrieval. In *International Computer Music Conference*, 2008.

[188] M. Yoshie, K. Kudo, and T. Ohtsuki. Effects of psychological stress on state anxiety, electromyographic activity, and arpeggio performance in pianists. *Medical Problems of Performing Artists*, 23(3):120, 2008.

[189] D. Young. *A Methodology for Investigation of Bowed String Performance Through Measurement of Violin Bowing Technique*. PhD thesis, MIT, 2007.

[190] V. Young, K. Burwell, and D. Pickup. Areas of study and teaching strategies in instrumental teaching: a case study research project. *Music Education Research*, 5(2), 2003.

Erklärung[1]

Hiermit erkläre ich, die vorgelegte Arbeit zur Erlangung des akademischen Grades "Dr.-Ing." mit dem Titel "Sensor-Based Feedback for Piano Pedagogy", selbständig und ausschließlich unter Verwendung der angegebenen Hilfsmittel erstellt zu haben. Ich habe bisher noch keinen Promotionsversuch unternommen.

Darmstadt, den 20. Mai 2011 Aristotelis Hadjakos

[1]Gemäß §9 Abs. 1 der Promotionsordnung der TU Darmstadt

Wissenschaftlicher Werdegang des Verfassers[2]

Aristotelis Hadjakos (geb. 1979 in Frankfurt a. M.) studierte Informatik an der Technischen Universität Darmstadt (Abschluss: Dipl.-Inform.). Parallel dazu studierte er an der Hochschule für Musik und Darstellende Kunst in Frankfurt Instrumental- und Gesangspädagogik (Klavier) in der Klavierklasse von Prof. Joachim Volkmann (Abschluss: Dipl.-Musiklehrer). Seit Februar 2007 ist Aristotelis Hadjakos Promotionsstudent unter der Betreuung von Prof. Dr. Max Mühlhäuser. Aristotelis Hadjakos war im Rahmen seiner Promotion Stipendiat des Landes Hessen bis zum Februar 2011. Seither ist er wissenschaftlicher Mitarbeiter im Fachgebiet Telekooperation.

[2] Gemäß §20 Abs. 3 der Promotionsordnung der TU Darmstadt